الدكتـور
محمـد صبح
استاذ في كلية العلوم
قسم الرياضيات

الدكتـور
عـدنان عمورة
استاذ مساعد في كلية العلوم
قسم الرياضيات

الدكتـور
عـزات قاسم
استاذ مساعد في كلية العلوم
قسم الرياضيات



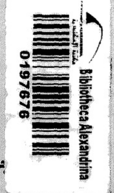
نظرية العينات

المدقـقون العلميـون

الأستاذ الدكتـور
... ابو حمدة

الأستاذ الدكتـور
انـور اللحـام

الأستاذ الدكتـور
محمـد صبـح



...ـامعة دمشـق

١٤٢٠ - ١٤٢١ هـ
١٩٩٩ - ٢٠٠٠ م

د. محمد صبح
أستاذ في كلية العلوم
قسم الرياضيات

د. عدنان عموره
أستاذ مساعد في كلية العلوم
قسم الرياضيات

د. عزات قاسم
أستاذ مساعد في كلية العلوم
قسم الرياضيات



المدققون العلميون

الأستاذ الدكتور
عبد الواحد أبو حمدة

الأستاذ الدكتور
أنور اللحام

الأستاذ الدكتور
محمد صبح

1420–1421 هـ

2000–2001 م

منشورات جامعة دمشق

نظرية العينات

مفردات المنهاج

- مبادىء نظرية العينات
- المعاينة العشوائية البسيطة
- المعاينة العشوائية الطبقية
- المعاينة العشوائية النمطية (المنتظمة)
- المعاينة العشوائية العنقودية
- تطبيق نظرية العينات في نظرية الاستقراء الإحصائي

مقدمة

إن علم الإحصاء (الرياضي والتطبيقي) يشكل اليوم أساساً قوياً لدراسة مختلف أنواع الظواهر والمسائل التي لها علاقة بالمجتمعات على مختلف أنواعها وتدخل العشوائية ضمن بنية هذا الظواهر والمجتمعات.

فعلم الإحصاء هو علم اتخاذ القرار في ضوء الحدس والتخمين. وإن من أهم عناصر الإحصاء التطبيقي هو المعانية العشوائية، حيث لها دور بارز في مختلف أعمال ونشاطات المراكز الإحصائية في العالم.

إن نظرية العينات تعد اليوم من أهم أجزاء العلوم الإحصائية المختلفة لما لها من دور كبير في الحصول على نتائج دقيقة وبكلفة قليلة.

إن كتابنا هذا يقدم مرجعاً علمياً أساسياً في علم نظرية العينات (المعانية العشوائية) وهو يخدم مقرر نظرية العينات لطلاب السنة الرابعة (شعبة الإحصاء) من قسم الرياضيات في كلية العلوم في جامعة دمشق. وهو يقدم أيضاً خدمات كبيرة لجميع المهتمين بالدراسات الإحصائية، لكي يتعرفوا على كيفية وطرائق اختيار العينات العشوائية من المجتمعات الإحصائية المدروسة، وأي العينات أفضل، وأي العينات تقدم دقة كبيرة وبأقل تكاليف ممكنة.

يتألف الكتاب من ستة فصول تعالج الأمور الأساسية في نظرية العينات وتطبيقاتها:

حيث نقدم في الفصل الأول المبادىء الأساسية في نظرية العينات ومفهوم هذه النظرية ومجالات تطبيقها وميزاتها وطرائق سحب العينات والإحصاءات الأساسية المستخدمة في هذه النظرية ومعايير جودة المقدّرات ويتخلل الفصل أمثلة محلولة وأمثلة غير محلولة تمكن الطالب من الإلمام بالأفكار النظرية والعملية لهذا الفصل.

والفصل الثاني يبحث في المعاينة العشوائية البسيطة، حيث يقدم هذا الفصل مفهوم المعاينة البسيطة وخواص التقديرات فيها ثم يدرس المعاينة العشوائية البسيطة مع الإعادة وبدون إعادة ثم يدرس الدقة وحجم العينة في هذه المعاينة. وتقدير النسبة ثم يـــدرس الارتباط بين خاصتين في المجتمع المدروس ويتضمن الفصل تمارين غير محلولة وأخـــرى محلولة لدعم أفكار هذا الفصل النظرية والعملية.

وفي الفصل الثالث يبحث في المعاينة العشوائية الطبقية حيث نستعرض مفهوم هذه المعاينة ودراسة التقديرات وخواصها في هذه المعاينة ، ثم ندرس المحاصّة المثلى، والدقــة النسبية في معاينة طبقية ومعاينة بسيطة ، ثم ندرس المعاينة الطبقية في حالة النســـب. ويتخلل الفصل تمارين محلولة وأخرى غير محلولة تدعم الأفكار النظرية والعملية لهـــذا الفصل.

وفي الفصل الرابع نقدم المعاينة النمطية ، حيث نصف هذه المعانية وندرس خواص التقديرات فيها وبعض المجتمعات الإحصائية الخاصة بها.

وفي الفصل الخامس ندرس المعاينة العشوائية العنقودية ، حيث ندرس حالة عناقيد متساوية الحجم وحالة عناقيد غير متساوية الحجـــم. ثم نســـتعرض فيـــها خـــواص التقديرات ، وندرس فيها أيضاً المعاينة العنقودية في حالة النسب.

والفصل السادس يبحث في تطبيق المعاينة العشوائية في نظرية الاستقراء الإحصائي حيث نستخدم المعاينة المدروسة في الفصول السابقة والمقدّرات في دراسة بناء مجـالات الثقة واختبار الفرضيات حول وسطاء المجتمعات الإحصائية المدروســـة بنـــاء علـــى معلومات (قيم إحصاءات) ناتجة عن عيّنات مسحوبة من تلك المجتمعات ويتخلل هذا الفصل تطبيقات محلولة وتطبيقات غير محلولة واسعة ليتمكن الدارس من اســـتطلاع مختلف المجالات والممكن تطبيق هذه النظرية فيها.

وأخيراً نقدم قائمة بأسماء المراجع الأساسية والمراجع الأخرى المتعلقة بهذه النظريـــة والتي يمكن الاستفادة منها للاستزاده والاطلاع.

ثم نقدم قائمة بالمصطلحات العلمية الإحصائية باللغتين الإنكليزية والعربية.

وإننا نضع هذا الكتاب بين أيدي زملائنا وطلابنا آملين أن نكون قد ســـاهمنا في إضافة ما يعوض شيئاً من النقص الذي تعانيه مكتباتنا العربية في مجال الكتب العلميـــة.

وسنكون شاكرين جزيل الشكر إلى كل من يقدم لنا ملاحظاته حول الكتاب ومضمونه كي نتمكن من معالجتها ، لكي يصبح الكتاب غنيا بموضوعاته ليعود بالنفع الكبير على أمتنا العربية

المؤلفون

الفصل الأول
مبادئ
نظرية العينات

1-1: مقدمة في نظرية العينات:

تعدّ نظرية العينات جزءاً لا يتجزأ من العلوم الإحصائية المختلفة وبخاصة بعد أن أصبح لهذه النظرية أسساً علمية رياضية، وهي من الناحية التطبيقية تتمتع بمكان متميز بين هذه العلوم حيث إن تطبيق هذه النظرية في البحوث العلمية يمكننا من الحصول على نتائج سريعة وبكلفة زهيدة نسبياً وإذا ما قيست بكلفة الإحصاء الشامل ، عدا عن أن نظرية العينات تعدّ الأسلوب الإحصائي الوحيد في كثير من حالات البحث العلمي كدراسة تكاثر الأسماك في البحر أو دراسة تطور البكتريا في وسط غذائي أو دراسة حركة الرمال في صحراء .....

وفي دراستنا لمقرر نظرية العينات سوف نستعرض مبادئ المعاينات العشوائية المختلفة وتبسيط الأسس الرياضية لنظرية العينات ثم حساب التقديرات المختلفة لوسطاء المجتمع ومقادير الأخطاء الناجمه بناء على العينات المأخوذة ، وبأسلوب يمكنّ كل مستثمر لهذه النظرية من استيعاب الأسس الرياضية لهذه النظرية وترجمة المبادئ والتصاميم والتقديرات إلى الواقع العملي وذلك في مختلف البحوث العلمية والتي يمكن أن تواجههم في المستقبل.

2.1 مفهوم نظرية العينات:

نظرية العينات هي دراسة للعلاقة القائمة بين المجتمع الإحصائي المطلوب دراسته والعينات المسحوبة منه . وهذه لها أهمية كبيرة في العديد من الأمور. فمثلاً: تكون ذات فائدة كبرى في تقدير الكميات غير المعلومة في المجتمع المدروس ، كمتوسط ذلك المجتمع أو تباينه أو انحرافه المعياري... والتي تدعى بوسطاء المجتمع وذلك من خلال المعلومات التي تقدمها العينة المسحوبة منه (كمتوسط العينة أو تباينها أو

إنحرافها المعياري) والتي تدعى بالإحصائيات المستخرجة من العينة. وأيضا تفيد نظرية العينات في تحديد ما إذا كانت الاختلافات المشاهدة بين عينتين تعود إلى المصادفة أو إلى اختلافات معنوية . وهذا يتم من خلال نظرية الاستقراء (الاستدلال) الإحصائي حيث نبين فترات (مجالات) ثقه ونختبر فرضيات حول وسطاء المجتمع معتمدين على معلومات تخص العينة ومستخدمين بذلك نظرية الاحتمالات ومن ثم نظرية القرار.

3.1 المعاينة العشوائية والأرقام العشوائية:

لضمان أن تكون الاستنتاجات المعتمدة في نظرية العينات والاستقراء الإحصائي سليمة فإن العينات المدروسة يجب أن تختار بحيث تكون ممثلة للمجتمع. وتدعى دراسة طرائق المعاينة والمشكلات المتصلة بها بتصميم التجارب. وإن إحدى طرائق الحصول على عينة ممثلة للمجتمع هي استخدام أسلوب ما يسمى المعاينة العشوائية ، وبناء عليه يكون لكل عنصر من عناصر المجتمع الفرصة نفسها في أن يكون من ضمن عناصر العينة . وإن أحد أبسط الأساليب في الحصول على عينة عشوائية هو إعطاء رقم لكل عنصر من عناصر المجتمع ونكتب هذه الأرقام على قطع ورقية ، صغيرة ومتماثلة ونضعها في كيس أو وعاء أو صندوق ونخلطها جيدا ثم نبدأ بسحب الأرقام من هذا الوعاء بمراعاة الخلط الجيد في كل عملية سحب. وهناك طرائق أخرى مثل استخدام جداول الأرقام العشوائية والتي صممت بشكل خاص لهذا الغرض.

4.1 المعاينة بإرجاع (مع الإعادة) أو بدون إرجاع (بدون إعادة):

عند عملية السحب يكون لنا الخيار في إعادة الرقم المسحوب أو عدم إعادته ففي حالة السحب مع الإعادة يمكن للعنصر أن يظهر في العينة عدة مرات وفي حالة السحب بدون إعادة فيظهر العنصر مرة واحدة فقط ومنه فلدينا:

معاينة عشوائية مع الإرجاع (مع الإعادة)

ومعاينة عشوائية بدون إرجاع (بدون إعادة).

5.1 توزيعات المعاينة:

لنعد كل العينات الممكنة ذات الحجم N والتي يمكن سحبها من مجتمع معين إما بإرجاع أو بدون إرجاع . ففي كل عينة يمكننا حساب إحصائية مثل: المتوسط الحسابي – التباين – الانحراف المعياري .... والذي سيختلف من عينة لأخرى.

وبهذه الطريقة نحصل على توزيع الإحصائية والذي يدعى بتوزيع المعاينة لهذه الإحصائية.

فمثلاً لو كانت الإحصائية المستخدمة هي المتوسط الحسابي للعينة ، فإن توزيعها يدعى بتوزيع العينة للمتوسطات. وبالطريقة نفسها يمكن أن نحصل على توزيعات المعاينة للتباين أو للانحراف المعياري أو للنسب أو ......

ولكل توزيع معاينة يمكننا حساب المتوسط الحسابي – التباين – الانحراف المعياري ..... وبذلك يمكن أن نتحدث عن المتوسط الحسابي لتوزيع المعاينة للمتوسطات أو تباين توزيع المعاينة للتباينات وغيرها.....

6.1 طرائق البحث الإحصائي:

إن كل البحوث الإحصائية تبدأ بمشاهدة ثم جمع المعلومات الإحصائية عن الموضوع المراد دراسته لمعالجتها وتحليلها وإن جميع هذه المعلومات الإحصائية يمكن أن يتم بطريقتين رئيستين:

1- **طريقة البحث الشامل:** وهنا يتناول الباحث جمع عناصر ووحدات المجتمع المدروس بدون استثناء أي منها وذلك بهدف الحصول على معلومات إحصائية شاملة ومن ثم إجراء التحاليل المنهجية اللازمة.

2- **طريقة البحث غير الشامل** : وهنا يتناول الباحث جزءاً معيناً ما أو نسبة معينة من عناصر المجتمع المدروس وذلك بهدف الحصول على معلومات إحصائية ودراستها ومن ثم تعميم نتائج هذه الدراسة على المجتمع المدروس ككل وهذه الطريقة تعرف باسم نظرية العينات.

ومنه ثم يمكننا استخدام معنى نظرية العينات حيث هي مجموعة الطرائق الرياضية والتنظيمية التي تساعدنا على إجراء البحوث الإحصائية غير الشاملة (على جزء من المجتمع المدروس) وذلك بهدف إيجاد الخصائص العامة للمجتمع المدروس وذلك بتعميم النتائج المستخلصة من هذه البحوث على ذلك المجتمع.

7.1 مميّزات نظرية العينات :

1- تختصر كثيراً من الوقت والجهد اللازمين لعمليات البحث الشامل وبالتالي ينتج اقتصاد بالكلفة.

2- تمكّن الباحثين من الحصول بسرعة على معلومات إحصائية مميّزة لعناصر المجتمع المدروس.

3- إن المعلومات الإحصائية المأخوذة بطريقة العينات هي أقل بكثير من مقابلتها والمأخوذة بطريقة البحث الشامل. مما يقلل من العمليات الحسابية والجهد والزمن اللازم لإنجازها.

4- تصحيح معلومات البحث الشامل عندما تكون نسبة الأخطاء فيها كبيرة وتفيدنا في معرفة الدقة المتوفرة في معلومات هذا البحث.

5- تبدو في بعض الحالات أن طريقة العينات هي الطريقة الرئيسة التي يمكن استخدامها لتعذّر أو استحالة إجراء البحث الشامل (كدراسة تطور البكتريا على وسط غذائي معين).

8.1 بعض مجالات تطبيق نظرية العينات:

1- في دراسة جودة إنتاج مصنع معين.

2- دراسة تقلّبات الأسعار في أسواق البورصة.

3- دراسة فعالية دواء جديد في علاج مرض معين.

4- دراسة التسويق وحركة المخازن وحركة المطارات والموانىء والمراكز الهاتفية...

5- دراسة ميزانيّات (أسر – منشآت – دول) وطريقة توزيعها أو صرفها.

6- دراسة توزع الأجور للعاملين في قطاع معين.

7- دراسة التجارب الزراعية (فعالية سماد معين – تطوير نبات معين ...).

8- دراسة حركة رأس المال (دخل – خرج – أرباح...).

9.1 تعاريف أولية في الإحصاء:

1- **المجتمع** : يقصد بالمجتمع كل الوحدات الأساسية التي تتضمنها المادة المدروسة والتي سنختار منها عينة ما لدراستها. ونرمز لحجمه بـ N .

2- **الإطار**: يقصد بالإطار جملة رموز عناصر المجتمع (ترقيم –تشفير – ....) مدرجة في قائمة أو موضوعة في جهاز معين على شكل بطاقات أو خرائط أو أرقام وذلك لإجراء السحب من هذا الجهاز.

3- **حجم العينة** : هو عدد عناصر العيّنة المسحوبة من المجتمع المدروس ونرمز له بـ n.

4- **وحدة المعاينة** : هي وحدة اصطلاحية غير قابلة للتجزئه تكون الأساس في عمليات سحب العيّنات ودراستها. ووحدة المعاينة التي تصلح لتجربه ليس من الضروري أن تصلح لتجربة أخرى.

مثل: (طول – حجم – وزن – أفراد – أنواع – ......)

5- **المعاينة:** هي عملية اختيار عدد ما من واحدات الإطار. وجملة العناصر المختارة تدعى عيّنة وعدد عناصرها يدعى بحجم العينة.

10.1 الخطوات الأساسية لتصميم العينة:

1- تحديد المشكلة والهدف المراد دراسته.

2- تعريف وتحديد المجتمع المراد معاينته وتحديد عناصره وتسمية وحدة المعاينة التي سيتناولها البحث وتحديد الفترة الزمنية المراد فيها إجراء هذه الدراسة والتأكد من خاصة التجانس.

3- تحديد المعلومات المطلوبة لإجراء الدراسة واستخراج النتائج.

4- تحديد طريقة جمع المعلومات : (قياس مباشر – اتصال مباشر – بريد – هاتف..).

5- تحديد الإطار الذي يغطي كل واحدات المجتمع.

6- تحديد حجم العينة وتكاليفها.

7- إجراء اختبار أولي للبحث عن بعض النواحي التي قد تكون غامضة في البدء وذلك لإدخالها في موضع الدراسة.

8- تلخيص وتحليل المعلومات وذلك بتبويب هذه المعلومات ومن ثم دراستها وتحليلها.

9- تقدير وسطاء المجتمع بوساطة العينة .

11.1 أنواع المعاينة:

1- **المعاينة العمدية:** يتعمد الباحث هنا اختيار عناصر معينة لإدخالها في العينة وهي حسب رأيه تمثل المجتمع المدروس تمثيلا جيدا وتستخدم في مجالات دراسة الرأي

العام – انتخابات معينة – أغراض دعائية. وهذه المعاينة تتضمن تحيزا ملحوظا لكـــن ممكن أن يكون أحيانا أقل من الخطأ العشوائي الناتج عن طريقة المعاينة العشـــوائية. وبكل الأحوال هذه المعاينة ليست بشكل عام جيدة.

2- **المعاينة العشوائية** : ومن أشهر أنواعها:

a- المعاينة العشوائية البسيطة

b- المعاينة الطبقية

c- المعاينة العنقوية.

d- المعاينة المنتظمة (الميكانيكية).

e- المعاينة التبديلية.

12.1 الشروط الأساسية للمعاينة العشوائية:

1- أن تكون عناصر المجتمع الكلي التي يراد أخذ العينة منه متجانسة من حيـــث طبيعتها ونوعية الدراسة المراد إجراؤها على هذه العناصر.

2- أن تكون عناصر المجتمع مستقلة عن بعضها بعضا أي أن انتقاء أي عنصر مـــن عناصر المجتمع لا يرتبط بسحب أو عدم سحب أي عنصر آخر.

3- أن يكون احتمال انتقاء أي عنصر من عناصر المجتمع الكلي معروفـــا لـــدى الباحثين أو يمكن حسابه.

4- أن يتم انتقاء عناصر العينة من المجتمع الكلي بدون تحيز أي أن يتصف الانتقـــاء بالعشوائية.

فالعينات المتصفة بهذه الشروط تدعى بالعينات العشوائية.

13.1 طرائق سحب العينات:

1- **بوساطة الكيس (الإطار)**: تقتضي هذه الطريقة أن نرقم جميع واحدات المجتمع بأرقام متسلسلة ونأخذ نسخه عن هذه الأرقام ونطويها ونضعها في كيس كبـــير ثم نخلطها جيدا وبعدها نقوم بالاختيار العشوائي بسحب أحد هذه الأرقـــام ونفتحـــة ونسجل الرقم المسحوب في قائمة عناصر العينة ثم نكرر العمل نفسه بعدد من المـــرات يساوي عدد عناصر العينة المطلوبة، ولكن لا بد لنا من أن نخلـــط الأرقـــام الـــتي في الكيس خلطا جيدا قبل كل عملية سحب.

2- **بوساطة الدواليب**: تقتضي هذه الطريقة أن نرقم جميع عناصر المجتمع بأرقام متتالية ثم نجهّز عدداً من الدواليب المتماثلة بمقدار عدد مراتب العدد N ونتأكد من أن هذه الدواليب ذات توازن تام وتعمل بصورة جيدة ونقسم محيط كل دولاب إلى عشرة أقسام متساوية نلصق على كل منها الأرقام 0,1,2,3,.......,9 وبعدها نثبّت سهماً خارج محيط هذه الدواليب للدلالة على الأعداد في عملية السحب. أما عملية الحصول على أحد الأرقام العشوائية فيتم بتدوير تلك الدواليب بقوة شديدة وبآن واحد ونتركها حتى تقف تماماً فيكون العدد الذي يشير إليه السهم على جميع هذه الدواليب هو العدد المسحوب ولذلك نسجله في قائمة واحدات العينة المسحوبة ثم نعيد عملية التدوير هذه حتى نحصل على كمية الأعداد اللازمة للحصول على العينة المطلوبة.

3- **بوساطة جداول الأرقام العشوائية**: وهي الأرقام العشرية من 0 إلى 9 مسجلة في جداول خاصة يأتي ترتيب هذه الأرقام فيها بطريقة عشوائية بحيث يكون احتمال سحب أي من هذه الأرقام في هذا الجدول معلوماً ومساوياً سحب أي رقم آخر.

ويتم تشكيل أي عدد بالطريقة التالية: نغمض أعيننا ونضع إصبعنا على الجدول فتقع على رقم ما نعده رقم الآحاد ونضع إصبعنا ثانية ، فتقع على رقم آخر نعده رقم العشرات.. وهكذا نكرر العملية حتى نشكلّ العدد الذي مرتبته تساوي مرتبة العدد (عدد واحدات المجتمع) وكما هو ملاحظ فإن تطبيق هذه الطريقة في سحب الأرقام العشوائية لا يحتاج إلى تنظيم إطار لواحدات المجتمع أيضاً.

مثال:

لنفترض مجتمعاً مؤلفاً من N=5000 عنصر معاينة ونريد سحب عينّة منه بحجم n=400 عنصر بوساطة الجداول العشوائية، نبدأ أولاً بترقيم واحدات هذا المجتمع بالأعداد المتتالية من 1 إلى 5000. وبما أن عدد واحدات المجتمع يتكّون من أربع مراتب من الأرقام الموجودة في جداول الأرقام العشوائية لذلك نغمض أعيننا ونضع إصبعنا على الجدول فنقع على رقم ما نعده رقماً لمرتبة الآحاد، ثم نضع إصبعنا مرة ثانية فتقع على رقم جديد نعده رقماً لمرتبة العشرات.. وهكذا حتى نحصل على رقم رابع نعده رقماً لمرتبة الألوف وبذلك نكون قد شكلّنا رقماً مؤلفاً من أربع مراتب ، فإذا وجد

في المجتمع واحده ما تحمل العدد نفسه فإننا نعمد إلى تسجيل هذا العدد في جـــدول خاص يدعى بجدول الأعداد المختارة وإلا فنهمل هذا العدد لنعيد هذه العملية عـــدداً من المرات يساوي كمية الأعداد المراد سحبها لتشكيل العيّنة المطلوبة.

فمثلاً إذا حصلنا على 0037 فهذا يعني أن الواحدة 37 من المجتمع هي عنصـــر في العينة المطلوبة.

وإذا حصلنا على 4302 فهذا يعني أن الواحدة 4302 من المجتمع هي عنصر آخـر في العينة المطلوبة.

وإذا حصلنا على 5627 فهذا يعني أن ذلك العدد غير موجود بين واحدات المجتمع وبالتالي نلغي عملية السحب هذه.

وفي حالة السحب بدون إعادة فإننا في حالة حصولنا على عدد مـــا كنّـــا قـــد سجلنّاه في جدول الأعداد المختارة فإننا نهمل هذا العدد ونعد هذه السحبة ملغيّة.

4- **بوساطة طريقة مونتي – كارلو**: تستخدم هذه الطريقة عندما يكون حجـــم العينة المراد سحبها كبيراً جداً وهي تعتمد على توليد أرقام شبه عشـــوائية وذلـــك بوساطة استخدام علاقة رياضية تولدّ هذه الأرقام من بعضها بعضاً وتأخذ هذه العلاقة الشكل التالي:

$$X_{i+1} = g(X_i) \qquad (1)$$

حيث X_i هو الرقم الذي جرى سحبه في عملية السحب ذات الرقم i

و X_{i+1} هو الرقم الذي يجري سحبه في عملية السحب ذات الرقم $i+1$

و g دالة رياضية يجب تعيينها

و X_0 يجب تحديده أو فرضه مسبقاً من قبل الباحث.

وخوارزمية مونتي – كارلو تكون كمايلي :

1- نرّقم واحدات المجتمع متسلسلة من 1 إلى N ونحدّد حجم العينة المراد سحبها.

2- نحدد احتمال سحب كل واحدة من واحدات ذلك المجتمع حيث إنه إذا كـــان السحب مع الإعادة فإن احتمال سحب كل واحدة منها يساوي $\frac{1}{N}$.

3- نقسم المجال [0,1] إلى N مجالاً متساوياً ونرقم هذه المجالات بالأرقام المتسلسلة من 1 إلى N. فإن طول كل مجال من هذه المجالات هو $\frac{1}{N}$ أي مساوياً لاحتمال سحب أي واحدة من واحدات المجتمع وتكون إحداثيات نقاط التقسيم هي على التوالي :

$$0,\frac{1}{N},\frac{2}{N},\frac{3}{N},\frac{4}{N},\ldots\ldots,\frac{N-1}{N},\frac{N}{N}=1 \qquad (2)$$

4- ليكن V متغيراً عشوائياً منتظم التوزيع على المجال [0, 1]

ويكون: (احتمال سحب أي عنصر من عناصر المجتمع) = $\frac{1}{N}$ = $P\,[V=x]$:

وبالتالي إذا قمنا بسحب عدد من قيم المتغير العشوائي وحددنا المجالات التي تقع فيها هذه القيم فإننا نكون قد حدّدنا وباحتمال قدره $\frac{1}{N}$ جملة من عناصر المجتمع والتي أرقامها تقابل هذه المجالات ونكون بالتالي قد حصلنا على عدد من عناصر المجتمع بطريقة عشوائية تتميّز بأن احتمال سحب كل عنصر من عناصر المجتمع متساوٍ ويساوي $\frac{1}{N}$.

ولكن سحب قيم المتغير العشوائي المختلفة يعد من أحد اختصاصات طريقة مونتي كارلو والتي تعتمد عدة طرائق لسحب هذه القيم وأهمها الطريقتان التاليتان:

a- بوساطة استخدام جداول الأرقام العشوائية:

هنا على سبيل المثال يتم سحب قيم المتغير العشوائي لعدد مؤلفة من أربعة مراتب ثم يقسم هذا العدد على 10000 وبذلك نحصل على كسر عادي يتلاءم مع قيم المتغير V ونكرر هذه العملية عدداً من المرات يساوي حجم العينة المطلوبة.

b- علاقة نيمان :

وهي من أشهر العلاقات الرياضية المستخدمة في توليد قيم المتغير العشوائي V وهي من الشكل

$$V_{i+1}=D[10^{-2k}\,y(10^{-2k}\,V_i^2)] \qquad (3)$$

حيث V_0 عدد كسري مفروض مسبقاً ولا يحوي الصفر في أي من مرتباته.

$2K$ هي مركبات قيم V ما بعد الفاصلة

y ترمز للجزء الصحيح من العدد الذي تؤثر عليه

D ترمز للجزء العشري من العدد الذي تؤثر عليه.

فإذا أردنا على سبيل المثال الحصول على قيم V ذات أربع مراتب ما بعد الفاصلة ، فإن العلاقة (3) تأخذ الشكل التالي :

$$V_{i+1} = D[10^{-4}\, y(10^{6}\, V_i^{2})] \qquad (4)$$

حيث

$$(2K = 4 \Rightarrow K = 2)$$

والعلاقة الأخيرة تعني بأن نأخذ المراتب الأربع الوسطى لمربع العدد V_i ولذلك تدعى هذه العلاقة بعلاقة وسط مربعات الأعداد.

5- نقارن V_{i+1} الحاصلة من الخطوة (4) مع إحداثيات المجالات المعرفه في الخطوة السابقة (3) حيث لا بد من وقوعها في أحد هذه المجالات فإذا كانت:

$$V_{i+1} \in [\frac{J-1}{N}, \frac{J}{N}] \Leftarrow \frac{J-1}{N} < V_{i+1} \leq \frac{J}{N}$$

والذي يقابل واحدة المجتمع ذات الرقم J ، لذلك فإننا نعد الوحدة ذات الرقم J من واحدات المجتمع هي الوحدة المختارة لتكون إحدى واحدات العينة المراد دراستها.

6- نعود إلى الخطوة (4) ونكرر العمليات حتى نحصل على العدد المطلوب سحبه لتشكيل العينة المطلوبة.

لكن علاقة نيمان ليست ذات فعالية جيدة وتولد أعدادا شبه عشوائية أي ليست عشوائية تماما ومتحيزة لصالح الأرقام الصغيرة وتشترط أن لا يحتوي V_0 على الصفر في أي من مرتباته وهذا بالإضافة إلى أنه يمكن أنه تتوقف هذه الخوارزمية عند عدد ثابت ما أو انعدام الأرقام. لذلك هناك علاقات أخرى في توليد الأعداد شبه العشوائية لكن لها أيضا سلبيات أخرى.

مثال: ليكن لدينا مجتمع مؤلف من 200 عنصر ونريد سحب عينة بحجم 5 عناصر.

الحل: نشكل النقاط التالية:

$$0,\frac{1}{200},\frac{2}{200},\frac{3}{200},\ldots,\frac{198}{200},\frac{199}{200},1$$

ولنفترض أن $V_0=0.5137$ حيث $2K=4$ أو $K=2$

$V_1=D[10^{-4}\,y(10^6\ V_0^2)]$

$=D[10^{-4}\,y(10^6.(0.26388769))]$

$V_1=0.3887$

بالمقارنة مع نقاط التقسيم نجد أن V_1 تقع في

$$\frac{77}{200}<V_1<\frac{78}{200}\Rightarrow V_1\in[\frac{77}{200},\frac{78}{200}]$$

فيكون العنصر $J_1=78$ هو العنصر الذي نعنيه بعملية السحب ويجب أن يدخل في واحدات العينة المسحوبة وبملاحظة ما جرى نجد أن: $J-1<NV_{i+1}\leq J$

ومنه NV_{i+1} يحتوي جزءا صحيحا وجزءا كسريا والعدد J هو أول عدد صحيح يلي العدد NV_{i+1} ومنه يتم اختياره بهذه الوسيلة.

نتابع العملية الآن للحصول على الأرقام الأخرى في العينة.

$V_1^2=0.15108769\Rightarrow V_2=0.1087\Rightarrow NV_2=21.74\Rightarrow \boxed{J_2=22}$

$V_2^2=0.01181569\Rightarrow \boxed{V_3=0.1815}\Rightarrow NV_3=36.3\Rightarrow \boxed{J_3=37}$

$V_3^2=0.03294225\Rightarrow \boxed{V_4=0.2942}\Rightarrow NV_4=58.84\Rightarrow \boxed{J_4=59}$

$V_4^2=0.08655364\Rightarrow \boxed{V_5=0.6553}\Rightarrow NV_5=(200)(0.6553)=131.06\Rightarrow \boxed{J_5=132}$

تطبيق : استمر في سحب عيّنة من الحجم 10 في هذا المثال.

ملاحظة:

إذا كان NV_{i+1} عدداً صحيحاً فهو J فعلاً ولا داعي للتقريب عندئذ.

ملاحظة: من المثال السابق وللتأكد من عدم فعالية علاقة نيمـــان ، نســـتمر في توليد الأرقام العشوائية حيث نجد : أن

$V_{40} = 0.1800$
$V_{41} = 0.2400$
$V_{42} = 0.7600 \Rightarrow J_{42} = 152$
$V_{43} = 0.7600 \Rightarrow J_{43} = 152$
$V_{44} = 0.7600 \Rightarrow J_{44} = 152$
$V_{45} = 0.7600 \Rightarrow J_{45} = 152$

وهنا نلاحظ أنه يمكننا أن نحصل على 42 رقماً مختلفاً وبعدها تبدأ العلاقة بإعطـاء رقم ثابت يقابل وحدة واحدة من واحدات المجتمع وبالتالي فلو أردنـــا ســـحب 100 واحدة من المجتمع باستخدام علاقة نيمان فسنحصل على 42 واحدة مختلفة ثم ستكرر إحدى واحدات المجتمع 58 مرة وهذا ما يجعل عملية السحب غـــير عشـــوائية. وإن مشكلة توليد الأرقام شبه العشوائية مازالت قيد البحث والدراسة.

c- علاقة توليد أرقام صحيحة:

بما أننا في نظرية العينات نهتم فقط بالحصول على الأرقام الصحيحــــة بصـــورة عشوائية لذلك فإننا سنحاول إيجاد علاقة رياضية لتوليد الأرقام الصحيحـــة ونـــأخذ الشكل :

$$n_i = y[10^k D[10^{-k-1}(23 i n_{i-1} + \zeta)]] \qquad (8)$$

حيث : $i = 1, 2, \ldots, h$ يمثل الرقم التسلسلي لعمليات السحب و:

n_0 عدد صحيح مفروض لا يحتوي الصفر ومرتباته من عدد مراتب العدد N.

K عدد مرتبات N فردية كانت أم زوجية.

ζ عدد ثابت مفروض لا يحتوي الصفر أيضاً وعدد مراتبه تساوي عدد مراتـــب العدد N على الأقل مهمته تصحيح الأرقام المولدّة بشكل دائم.

Y وD يرمزان للجزء الصحيح والكسري من العدد اللذين يؤثران فيه وبناء علـــى العلاقة (8) يمكننا أن نلخص خطوات الخوارزمية في سحب الأرقـــام الصحيحـــة ، كمايلي :

(1) نرقم عناصر المجتمع من 1 إلى N ونحدد حجم العينة المراد سحبها.

(2) نحدّد العددين ζ و n_0 وفقاً لشروط تحديديهما.

(3) نجري عملية التوليد وفقاً للعلاقة (8) أعلاه.

(4) نقارن الرقم n_i الذي حصلنا عليه من الخطوة (3) مع الرقم N فإذا كان $n_i \leq N$ عندها تسجل الرقم n_i في لائحة واحدات العينة المسحوبة ونعد الواحدة ذات الرقم n_i هي الواحدة المسحوبة ونفرزها عن المجتمع.

وإذا كان $n_i > N$ هنا يكون مرفوض ونعود لـ (3)

ونكرر هذه العمليات حتى نحصل على العينة المطلوبة. من عيوب هذه الخوارزمية هي أن العلاقة (8) تعد معقدة حسابياً بالنسبة للعلاقات الأخرى، وأن هناك العديد من الحسابات ستجرى بدون فائدة نتيجة لمقارنة n_i مع N مما يعد ضياعاً لوقت الحاسوب. ولكن مع هذا فإن العلاقة (8) وخوارزميته تعطيان نتائج جيدة عند توليد الأرقام العشوائية الصحيحة بوساطتها. والتجارب الأولية التي أجريت على هذه العلاقة دلت على أن الأرقام الناتجة تتصف بعشوائية كافية وموزعة بانتظام تقريباً على المجالات المتساوية.

مثال: ليكن مجتمع مؤلف من 5000 عنصر ونريد سحب عينة منه بحجم 50 عنصراً معتمدين على خوارزمية توليد الأعداد الصحيحة.

حيث: $\zeta = 2589$ و $n_0 = 3476$

الحل : لدينا K= 4 ونجد أن

$$n_i = y[10^k D[10^{-5}(23)(1)\, n_0 + \zeta)]]$$
$$= y[10^4 D[10^{-5}(23)(3476) + 2589)]] =$$
$$n_1 = y[10^k D[0.82537]] = y[10^4 (0.82537)]$$
$$= y[8253.7] = 8253 \Rightarrow \boxed{n_1 = 8253}$$

ولكن $n_1 > 5000$ ومنه لا نسجله في لائحة الأرقام المسحوبة، بل نستخدمه لتوليد الأرقام المتتالية وبالطريقة نفسها نجد

$n_2 = 8222$ مرفوض , $n_3 = 6990$ مرفوض , $n_4 = 4566$ (مقبول)

$n_5=2767$ (مقبول) , $n_6=8443$ (مرفوض)

$n_7=5932$ مرفوض , $n_8=6191$ مرفوض , $n_9=8414$, $n_{10}=3734$

ونستمر في 98 مره سحب للحصول على عينة من 50 عنصرا

$$n_{94}=\boxed{3607}, n_{95}=8388, n_{96}=\boxed{2329}$$

$$n_{97}=\boxed{9858}\ , n_{98}=\boxed{2019}$$

وبما أننا احتجنا 98 مرة سحب لتشكيل عينة عشوائية حجمها 50 فهذا لا يدل على عيب في العلاقة بل لأن N = 5000 تشكل تقريبا نصف الـ 9999 وبهذا احتجنا إلى عمليات سحب تساوي ضعف الحجم المطلوب وهذا ما يدل على أن الأرقام موزعة تقريبا منتظما في المجالين [5000.9999] , [0.5000]

14-1 المعاينة وتصنيفها:

إن أسلوب المعاينة يستخدم في مختلف أنواع الدراسات الاقتصادية – الصناعية – الزراعية – التجارية – الصحية – التربوية والنفسية – الطبية – وتصنف مسوحات العينة إلى نوعين:

1- **مسح وصفي**: يهدف إلى الحصول على معلومات معينة حول مجموعات ضخمة مثل عدد الرجال والنساء والأطفال الذين يشاهدون برنامجا تلفزيونيا معينا.

2- **مسح تحليلي**: يهدف إلى المقارنة بين مجموعات جزئية مختلفة من المجتمع بغية اكتشاف ما إذا كانت هناك فروق معينة ولصياغة أو التحقق من فرضيات تتعلق بأسباب هذه الفروق وهناك العديد من مسوحات البيانات التي تخدم كلا من الهدفين وبالتالي تهدف نظرية المعاينة إلى جعل المعانية أكثر كفاءة حيث تحاول تطوير طرائق اختيار عناصر العينة وتطوير طرائق الوصول إلى التقديرات التي ستنتجها العينة بأقل كلفة ممكنة وبدقة عالية أو دقة محددة بأقل كلفة ممكنة وبما يتعلق بالدقة هناك أسباب جيدة لفرض أنه التوزيع الاحتمالي للمقدرات الناتجة على وجه التقريب طبيعية وبالتالي فإذا علمنا المتوسط والإنحراف المعياري (أو التباين) نكون قد عرفنا شكل دالة

التوزيع التكراري بكاملها. ومنه نظرية العينات تهتم بجزء كبير منها في إيجاد علاقــات خاصة بالمتوسطات والتباينات.

15-1 الإحصائيات المستخدمة في نظرية العينات:

سندرس أهم المؤشرات الإحصائية (الإحصائيات) المستخدمة وذلــك في حالــة السحب مع الإعادة وبدون إعادة وأهمها:

1- **عدد واحدات المجتمع الكلي:** المدروس ونرمز له بـــ N

2- **عدد واحدات العينة المسحوبة من المجتمع** : ونرمز له بـــ n وندعوه بحجــم العينة.

3- **إجمالي خاصة ما في المجتمع** : وهو بالتعريف مجموع قيم خاصة مــا لواحدات المجتمع ونرمز له بـــ Y فإذا كانت قيم خاصة ما في واحدات المجتمع هي:

$$y_1, y_2, \ldots\ldots, y_N \Rightarrow Y = \sum_{i=1}^{N} y_i = y_1 + y_2 + \ldots\ldots + y_N$$

4- **إجمالي خاصة ما في العينة** : وهو بالتعريف مجموع قيمّ خاصة ما لواحـــدات العينة ذات الحجم n فإذا كانت قيم خاصة ما في العينة هـــي $x_1, x_2, \ldots\ldots, x_n$ فـــإن مجموع هذه القيم X

$$X = x_1 + x_2 + \ldots\ldots + x_n \quad = \sum_{i=1}^{n} x_1$$

حيث هناك y_i من المجتمع تقابل x_i في العينة من أجل $i = 1, 2, \ldots\ldots, n$.

5- **المتوسط الحسابي لقيم خاصة ما في المجتمع:** من المعروف أن المتوسط الحسابي يعرّف بالعلاقة :

$$\overline{Y} = \frac{Y}{N} = \frac{\sum_{i=1}^{N} y_i}{N} \Rightarrow Y = N.\overline{Y}$$

6- **المتوسط الحسابي لقيم خاصة ما في العينة ذات الحجم n:**

$$\overline{X} = \frac{X}{n} = \frac{\sum_{i=1}^{n} x_i}{N} \Rightarrow X = n.\overline{X}$$

7- عدد العينات ذات الحجم n: والممكن تشكيلها من مجتمع مؤلـــف مـــن N عنصراً، تميّز هنا نوعين من العينات:

a- السحب مع الإعادة

$$C_{N+n-1}^{n}=\binom{N+n-1}{n}=\frac{(N+n-1)!}{n!(N-1)!}$$

b- السحب بدون إعادة:

$$C_{N}^{n}=\binom{N}{n}=\frac{N!}{n!(N-n)!}$$

والعينات متوسط حسابي خاص بها وبالتالي $\overline{X}$ يعد متغيراً عشوائياً وبالتالي يجـدر بنا أن نبحث عن التوزيع الاحتمالي الذي يخضع له هذا المتغير العشوائي.

وإن شكل توزيع $\overline{X}$ مرتبط بشكل توزيع القيم y_i في المجتمع ومن أجل N كبيرة فإن القيم y_i تتوزع على الغالب توزعاً طبيعياً

من الشكل:

$$\phi(y)=\frac{1}{\sigma\sqrt{2\pi}}.e^{\frac{-(y-\overline{Y})^2}{2\sigma^2}}$$

حيث σ^2 هو تباين y_i في المجتمع و $\overline{Y}$ متوسط الحسابي. ومنه $\overline{X}$ تتوزع أيضـاً طبيعياً من الشكل:

$$\phi(\overline{X})=\frac{1}{\sigma_{\overline{X}}.\sqrt{2\pi}}.e^{\frac{-(\overline{X}-\overline{Y})^2}{2\sigma_{\overline{X}}^2}}$$

حيث $\sigma_{\overline{X}}^2$ هو بتاين متوسطات العيّنات عن $\overline{Y}$ ويمكننا أن نعود للتوزيع الطبيعـــي بتطبيق مبرهنة النهاية المركزية من أجل $n>30$.

8- احتمال انتقاء عينة ذات حجم n من مجتمع مؤلف من N عنصـــراً: بعـــد معرفتنا بعدد العينات ذات الحجم n والممكن تشكيلها من N عنصراً فـــإن احتمـــال انتقاء أي من هذه العينات يساوي:

a- في حالة السحب بدون إعادة:

$$P_n = \frac{1}{\binom{N}{n}}$$

b- في حالة السحب مع الإعادة:

$$P_n = \frac{1}{\binom{N+n-1}{n}}$$

حيث فرضنا أن فرض اختيار أي من هذه العينات متساوية.

9- احتمال انتقاء عنصر معين A من المجتمع في السحبة i :

a- حالة السحب مع الإعادة:

$$P_A = \frac{1}{N}$$

b- حالة السحب بدون إعادة :

$$P_A(r) = \frac{1}{N-(r-1)} = \frac{1}{N-r+1}$$

لأن احتمال سحب العنصر A في المرة الأولى: $\frac{1}{N}$

واحتمال سحب العنصر A في المرة الثانية: $\frac{1}{N-1}$

واحتمال سحب العنصر A في المرة الثالثة: $\frac{1}{N-2}$

: : : : :

واحتمال سحب العنصر A في المرة r : $\frac{1}{N-(r-1)}$

10- احتمال وجود عنصر معين a في العينة ذات الحجم n: بما أن العينة من الحجم n فإن انتقاء هذا العنصر a يتمتع بـ n فرصة مستقلة (أي n عملية سحب). لذا فإن احتمال انتقاء العنصر a في العينة ذات الحجم n يساوي مجموع احتمالات انتقائه في كل عملية سحب ونجد أن:

a- حالة السحب مع الإعادة :

$$P_n(a) = \frac{1}{N} + \frac{1}{N} + \dots\dots\dots + \frac{1}{N} = \frac{n}{N}$$

b- حالة السحب بدون إعادة:

$$P_n(a)=\frac{1}{N}+\frac{1}{N-1}+\frac{1}{N-2}+......+\frac{1}{N-(n-1)}$$

11- التوقع الرياضي لمتوسطات العينات: نعلم أن التوقـــع الريـــاضي لمتغـــير عشوائي يعطى بالعلاقة:

$$E(Z)=\sum_{K=1}^{M} z_K P_K$$

حيث Z_K تمثل قيم المتغير العشوائي Z المختلفة والتي عددها M و P_K تمثـــل احتمالات ظهور تلك القيم (أي تكرارها النسبية). ومنه يكون توقـــع متوســـطات العينات:

a- حالة السحب مع الإعادة :

$$E(\overline{X})=\sum_{K=1}^{\binom{N+n-1}{n}} \overline{X}_K . \frac{1}{\binom{N+n-1}{n}}$$

b- حالة السحب بدون إعادة:

$$E(\overline{X})=\sum_{K=1}^{\binom{N}{n}} \overline{X}_K . \frac{1}{\binom{N}{n}}$$

حيث يكون المجموع مأخوذا على جميع العينات الممكنة في حالة السحب

12- تباين متوسطات العينات $\overline{X}$ عن متوسط المجتمع $\overline{Y}$:

$$V(\overline{X})=\sigma^2_{\overline{X}}=E(\overline{X}_K-\overline{Y})^2$$

a- حالة السحب بدون إعادة:

$$\sigma^2_{\overline{X}}=\sum_{K=1}^{\binom{N}{n}}(\overline{X}_K-\overline{Y})^2 \frac{1}{\binom{N}{n}}$$

b- حالة السحب مع الإعادة:

$$\sigma^2_{\overline{X}}=\sum_{K=1}^{\binom{N+n-1}{n}}(\overline{X}_K-\overline{Y})^2 . \frac{1}{\binom{N+n-1}{n}}$$

13- الانحراف المعياري لـ $\bar{X}$:

وهو الجذر التربيعي الموجب لـ $\sigma_{\bar{X}}^{2}$ (تباين متوسط العينة).

ملاحظة:

إن عدد العينات المتمايزة من الحجم n والمكن سحبها من مجتمع عدد عناصره N من الشكل:

(حالة السحب مع الإعادة) $|\Omega| = N^{n}$

14- توزيع المعاينة: للمتوسطات : لنفترض كل العينات الممكنة ذات الحجم n والمسحوبة من مجتمع عدد عناصره N حيث n<N ونرمز للمتوسط الحسابي لتوزيع المعاينة بـ $\mu_{\bar{X}}$ والإنحراف المعياري بـ $\sigma_{\bar{X}}$ ونرمز لمتوسط المجتمع بـ μ والإنحراف المعياري بـ σ فإن:

a- حالة السحب بدون إعادة :

$$\mu_{\bar{X}} = \mu \quad ; \quad \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}$$

b- وفي حالة السحب مع إعادة :

$$\mu_{\bar{X}} = \mu \quad ; \quad \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

وسنبرهن ذلك لاحقاً. وبكل الأحوال فإن توزيع المعاينة للمتوسطات يتوزع تقريباً طبيعياً بتوقع $\mu_{\bar{X}}$ وانحرافي معياري $\sigma_{\bar{X}}$ وذلك بصرف النظر عن المجتمع وكان حجم المجتمع ضعف العينة على الأقل.

15- ثابت التباين النسبي : وهو نسبة التباين لـ $\bar{X}$ على $\bar{X}^2$ أي:

$$C^2 = \frac{\sigma_{\bar{X}}^{2}}{\bar{X}^2}$$

سواء كان السحب مع الإعادة أم بدون إعادة حسب الحالة.

16- **مدى الثقة ومعامل الثقة:** عندما ندرس واحدات العينة ونحصل من خلالهـا على تقدير ما $\hat{a}$ لمقدار a في المجتمع فإن الأمر يتطلب منا الاهتمام بمقـدار الخطـأ المركب من جراء ذلك التقدير، وبما أن a نفسها هي مقدار مجهول فإن حساب مثـل ذلك الخطأ مباشرة يعد أمراً مستحيلاً ونتيجة لذلك نبحث عن المجال الذي يمكـن أن يقع فيه المقدار a باحتمال قدره β أي نبحث عن العدد الموجب ε_β الذي يحقـق المعادلة التالية:

$$P[|\hat{a}-a|<\varepsilon_\beta]=\beta$$

فإذا استطعنا أن نجد ε_β فإننا نقول إن الخطأ المرتكب والناتج عن تقدير a بوساطة $\hat{a}$ محصور في المجال

$[\hat{a}-\varepsilon_\beta, \hat{a}+\varepsilon_\beta]$ باحتمال قدره β.

وأن $a \in [\hat{a}-\varepsilon_\beta, \hat{a}+\varepsilon_\beta]$ باحتمال قدره β.

وتشير هنا إلى أن المقدار المجهول a هو مقدار ثابت وليس له أي صفة عشـوائية، بينما طول المجال السابق هو الذي يتصف بالعشوائية. فلإيجاد العدد ε_β نحتـاج إلى معرفة دالة التوزيع الاحتمالي للمقدار $(\hat{a}-a)$ فلذلك نفترض أن هذه الدالة معروفـة ونرمز لها بـ $F(x)$ ومنه

$$P[|\hat{a}-a|<\varepsilon_\beta]=F(\varepsilon_\beta)-F(-\varepsilon_\beta)$$

وبما أننا نرغب أن يكون ذلك الاحتمال مساويا لـ β المفروضة ، عندئذ:

$$\boxed{F(\varepsilon_\beta)-F(-\varepsilon_\beta)=\beta}$$

وبحل هذه المعادلة نحصل على قيمة ε_β المقابلة لقيمة β في دالة التوزيع $F(x)$ ومنه نسمي المجال:

$[\hat{a}-\varepsilon_\beta, \hat{a}+\varepsilon_\beta]$ بمجال الثقة حول a ونسمي الاحتمال β باحتمـال الثقـة (مستوى الثقة) لانتماء a للمجال السابق.

وكحالة خاصة وهي أن توزيع المعاينة يكون طبيعيا عندئذ

$$F(x)=\frac{1}{\sigma_{\hat{a}}.\sqrt{2\pi}}\int_{-\infty}^{x} e^{\frac{-(\hat{a}-a)^2}{2\sigma_{\hat{a}}^2}}\, d\hat{a}$$

ونأخذ $t=\frac{\hat{a}-a}{\sigma_{\hat{a}}}$ فإن دالة التوزيع المعيارية:

$$\phi(Z)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{Z} e^{\frac{-t^2}{2}}\, dt$$

ومنه يمكننا أن نحسب ε_β بالشكل التالي:

$$P[|\hat{a}-a|<\varepsilon_\beta]=P[\frac{|\hat{a}-a|}{\sigma_{\hat{a}}}<\frac{\varepsilon_\beta}{\sigma_{\hat{a}}}]$$

$$\Leftrightarrow P[|t|<\frac{\varepsilon_\beta}{\sigma_{\hat{a}}}]=\phi(\frac{\varepsilon_\beta}{\sigma_{\hat{a}}})-\phi(\frac{-\varepsilon_\beta}{\sigma_{\hat{a}}})$$

وكون $\phi(-Z)=1-\phi(Z)$ (خاصة التناظر في التوزيع الطبيعي)
عندئذ :

$$P[|\hat{a}-a|<\varepsilon_\beta]=2\phi(\frac{\varepsilon_\beta}{\sigma_{\hat{a}}})-1$$

وبأخذ المقدار الأخير مساوياً β فنحصل عندئذ:

$$2\phi\,(\frac{\varepsilon_\beta}{\sigma_{\hat{a}}})-1=\beta\Rightarrow\phi(\frac{\varepsilon_\beta}{\sigma_{\hat{a}}})=\frac{1+\beta}{2}$$

$$\Rightarrow\varepsilon_\beta=(\sigma_{\hat{a}})\phi^{-1}(\frac{1+\beta}{2})=Z_\beta.\sigma_{\hat{a}}$$

إن $\sigma_{\hat{a}}$ بحيث من قياسات العينة و Z_β تتمثل قيمة المتغير العشوائي المقابلة للقيمة الاحتمالية $\frac{1+\beta}{2}$ للدالة ϕ .

إن قيمة $\phi^{-1}(\frac{1+\beta}{2})$ يمكن الحصول عليها من الجدول التالي: (لبعض القيم) وحالة توزيع طبيعي.

β	0.90	0.94	0.95	0.96	0.98	0.99	0.9973	0.80
$Z_\beta = \phi^{-1}(\frac{1+\beta}{2})$	1.643	1.88	1.96	2.053	2.33	258	3	1.282

وبالتالي بعد إيجاد Z_β نشكل مجال الثقة

$$I_\beta = [\hat{a} - Z_\beta \sigma_{\hat{a}}, \hat{a} + Z_\beta \sigma_{\hat{a}}]$$

ذلك المجال الذي يغطي المقدار a باحتمال قدره β .

مثال: ليكن $\hat{a} = 10.73$ تقديرا لـ a وليكن $\sigma_{\hat{a}} = 0.0564$

والمطلوب إيجاد مجال الثقة الذي يغطي a بثقة $\beta = 0.80$

الحل: $a \in [\hat{a} - Z_\beta \sigma_{\hat{a}}, \hat{a} + Z_\beta \sigma_{\hat{a}}]$

$$a \in [10.73 - (1.282)(0.0564) , 10.73 + (1.282)(0.0564)]$$

$$a \in [10.71, 10.85]$$

17- الخطأ المطلق لتقدير ما: ليكن لدينا $\hat{a}_K$ تقديرا ما للمقدار a ، فإننا نعرف الخطأ المطلق للتقدير $\hat{a}$ بـ $\delta_K = |\hat{a} - a|$

18- الدقة : تعرف الدقة بأنها الحد الأعلى للخطأ المطلق ونرمز لها بالشكل التالي:

$$d = \sup \delta_K = \sup |\hat{a}_K - a|$$

وبالتالي

$$d \geq |\hat{a}_K - a|$$

19- الدقة النسبية : تعرف الدقة النسبية بـ

$$d = \frac{d}{\overline{X}} = \frac{Z_\beta \sigma_{\overline{X}}}{\overline{X}}$$

20- نسبه الواحدات التي تتصف بخاصة معينة:

كثيرا ما نتعرض في بحوث العينات إلى دراسة نسبة واحدات المجتمع التي تتصف بخاصة ما. فإذا رمزنا لعدد هذه الواحدات في المجتمع بـ A فإن نسبة هذه الواحدات في المجتمع تساوي

$$P=\frac{A}{N}\Rightarrow \phi=\frac{N-A}{N}=1-P$$ (بنسبة الواحدات التي لا تتصف بهذه الخاصة)

وإذا سحبنا عينة ذات حجم n ووجدنا فيها a عنصرا يتصف بتلك الخاصة. فإن نسبة وجود تلك العناصر في العينة.

$$p=\frac{a}{n}\Rightarrow q=\frac{n-a}{n}=1-p$$

أما قانون توزيع ظهور a عنصرا يتصفون بالخاصة المذكورة فيمكن أن نعالجة كما يلي:

من نظرية الاحتمالات نعلم أن احتمال حصولنا على عنصر يتصف بخاصة ما m مرة في حالة إجرائنا تجارب سحب عشوائية عددها n أي $m\leq n$ يعطي بالعلاقة

$$P(n,m)=\binom{n}{m}p^{m}q^{n-m}$$

حيث p يمثل احتمال ظهور العنصر الذي يتصف بتلك الخاصة وأن $q=1-p$.

ومن أجل $m=a$:

$$P(n,a)=\binom{n}{a}p^{a}q^{n-a}$$

و:

$$P(n,K\,K\leq a)=\sum_{K=0}^{a}\binom{n}{K}p^{K}q^{n-K}$$

16-1 معايير جودة التقدير:

وهي مهمة جدا في نظرية العينات وهذه المعايير هي:

a- عدم التحيز: نقول عن المقدار $\hat{a}_K$ إنه يعد تقديرا غير منحاز لـ a فيما إذا كان التوقع الرياضي لـ $\hat{a}_K$ يساوي a أي $E(\hat{a}_K)=a$ حيث K دليل يرمز إلى جميع التقديرات الممكنة المأخوذة من جميع العينات الممكنة ذات الحجم n. وإذا كان:

$E(\hat{a}_K)\neq a$ عندئذ $\hat{a}_K$ هو تقدير متحيز لـ a أي أنه لا يمثل المقـــدار a تمثيـــلا صحيحا.

مثال: إن $\overline{X}_K$ هو مقدر غير منحاز لـ $\overline{Y}$ لأن $E(\overline{X}_K)\neq\overline{Y}$

b- التماسك أو التراص أو الاتساق:

نقول عن المقدار $\hat{a}_K$ إنه يمثل تقديرا متماسكا للمقدار a فيمـــا إذا كـــان $\hat{a}_K$ يتقارب بالاحتمال من a نفسه وذلك عندما يزداد عدد عناصر العينة إلى اللانهايـة أو إلى N أي

$$\lim_{n\to N(\infty)} P(\hat{a}_K \to a)=1$$

حيث P ترمز إلى الاحتمال وذلك حسب قانون التوزيع لـ $\hat{a}$.

مثال:

إن $\overline{X}=\frac{\sum_{i=1}^{n}x_i}{n}$ و إن $\overline{Y}=\frac{\sum_{i=1}^{N}y_i}{N}$ وأن $\lim_{n\to N} \overline{X}=\overline{Y}$

وهذا يعني أن متوسط العينة يعد تقديرا متماسكا لمتوسط المجتمع لأن يحقق العلاقة السابقة.

ملاحظة:

لاحظنا أن المتوسط يمثل تقديرا متماسكا لمتوسط المجتمع وأيضا يمثل تقديرا غـــير متحيز لذلك المتوسط، فهذا لا يعني أن معياري عدم التحيز والتماســـك مرتبطـــان ببعضها بعضا بل على العكس إن معيار التماسك مستقل تماما عن معيار عدم التحـــيز ولا ينتج عنه بالضرورة والعكس بالعكس.

C- الفعالية : في أغلب الأحيان يمكن أن نجد لمقدار ما a عـــددا كبـــيرا مـــن التقديرات غير المتحيزة والمتماسكة وذلك حسب حجوم العينات المسحوبة بطريقـــة تصميم المعاينة. وعندها لا بد لنا من معيار آخر لاختيار أفضل تلك التقديرات ومـــن

$$Var(\hat{a}_k)=\frac{\sum_{K=1}^{M}(\hat{a}_k-a)^2}{M}=\text{mi}$$

مثال: إن متوسط العينة يعد تقديراً غير متحيز ومتماسكاً لمتوسط المجتمع ولكن تباينة هو $Var(\bar{X})=\frac{\sigma^2}{n}$ وهو أصغر من تباين أي تقدير آخر.

وتجدر الإشارة هنا إلى أنه يمكننا إيجاد تقدير متحيزّ وذي تباين أصغر لمقدار a يكون أفضل من تقدير غير متحيزّ وذو تباين كبير نسبياً .

d- التقدير الأمثلي: نقول عن التقدير $\hat{a}_k$ إنه أمثلي للمقدار a إذا حقق $\hat{a}_k$ المعايير الثلاثة السابقة: غير متحيزّ ومتماسك وذي تباين أصفري.

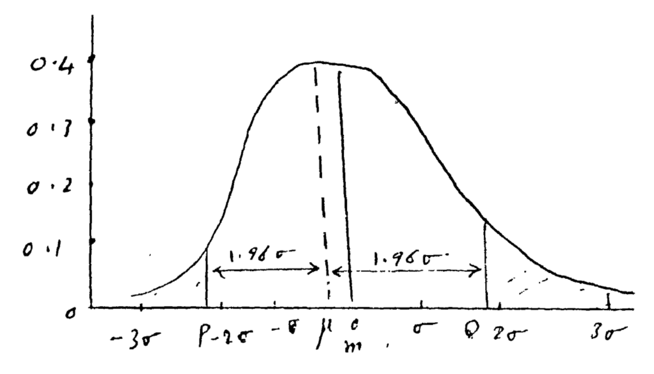
17-1 الانحياز وتأثيره: من الضروري في نظرية العينات أن نأخذ بالحسبان التقديرات المنحازة لسببين:

a- في بعض المسائل الأكثر شيوعا ، وبوجه خاص في تقديرات النسب ، نجد أن التقديرات المريحة والمناسبة هي تقديرات منحازة.

b- وفي حالة المعاينة الاحتمالية نجد أنه حتى مع التقديرات غير المنحازة يمكن أن تنتج أخطاء القياس وعدم الاستجابة انحيازا في الأعداد التي نستطيع حسابها من بيانات العينة.

مثال: إذا كان الأشخاص الذين رفضوا إجراء مقابلة كلهم تقريبا من المعارضين لنوع صرف الأموال العامة ، بينما ينقسم أولئك الذين أجروا المقابلة بالتساوي بين مؤيد ومعارض

لدراسة تأثير الانحياز ، نفترض أن التقدير $\hat{\mu}$ يتوزع طبيعيا حول المتوسط m يقع على مسافة B من المتوسط الحقيقي للمجتمع μ كما في الشكل :



(شكل يبيّن تأثير الانحياز في أخضاء عملية القياس) ومقدار الانحياز $B=m-\mu$.

لنفترض أننا لانعلم بوجود أي انحياز ، ولنحسب الانحراف المعياري σ للتوزيع التكراري الموافق للتقدير، وسيكون هذا بالطبع الانحراف المعياري حول المتوسط المفترض للتوزيع m وليس حول المتوسط الحقيقي μ. وكعبارة حول دقة التقدير نقول إن احتمال أن يتجاوز خطأ التقدير $\hat{\mu}$ الكمية 1.96σ هو 0.05 فقط.

وسنرى كيف يحرف وجود الانحياز هذا الاحتمال. ولتوضيح ذلك نحسب الاحتمال الصحيح لكون الخطأ المرتكب في التقدير أكبر من 1.96σ حيث نقيس الخطأ بدءاً من المتوسط الحقيقي μ ولا بد من تأمل كل من ذيلي التوزيع على حدة فبالنسبة للذيل الأعلى ، يكون أحتمال وجود خطأ بالزيادة أكبر من 1.96σ يمثل المساحة المظللة على يمين σ (في الشكل) وهذه المساه هي:

$$\frac{1}{\sigma\sqrt{2\pi}}\int_{\mu+1.96\sigma}^{\infty} e^{\frac{-(\hat{\mu}-m)^2}{2\sigma^2}}.d\hat{\mu}$$

وبأخذ $\hat{\mu}-m=\sigma\, t$ عندئذ يصبح الحد الأدنى للتكامل بدلالة t

$$\frac{\mu-m}{\sigma}+1.96 \quad ;t=\frac{\hat{\mu}-m}{\sigma} \Rightarrow t=1.96-\frac{B}{\sigma}$$

وتصبح المساحة من الشكل:

$$\frac{1}{\sqrt{2\pi}} \int_{1.96-\frac{B}{\sigma}}^{\infty} e\frac{-t^2}{2}.dt$$

وبصورة مماثلة فإن الذيل الأدنى أي المساحة المظللة على يسار P تساوي

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-1.96-\frac{B}{\sigma}} e\frac{-t^2}{2}.dt$$

ويتضح من شكل التكاملين أن مقدار الاضطراب يعتمد فقط على نسبة الانحياز إلى الانحراف المعياري والنتائج مبينة في الجدول التالي:

B/σ	$<-1.96\sigma$	$>1.96\sigma$	المجموع
0.02	0.0238	0.0262	0.0500
0.04	0.0228	0.0274	0.0502
0.06	0.0217	0.0287	0.0504
0.08	0.0207	0.07301	0.0508
0.10	0.0197	0.0314	0.0511
0.20	0.0154	0.0392	0.0546
0.40	0.0091	0.0594	0.0685
0.60	0.0052	0.0869	0.0921
0.80	0.0029	0.1230	0.1859
1.00	0.0015	0.1685	0.1700
1.50	0.0003	0.3228	0.3231

وفيما يتعلق بالاحتمال الكلي لوجود خطأ أكبر من 1.96σ ، نجد أن للانحياز تأثيرا ضعيفا شريطة أن يكون هذا الانحياز أقل من عشر الانحراف المعياري. وعندما يكون الانحياز $\beta=0.10\sigma$ فإن الاحتمال الكلي ويكون 0.0511 بدلا من 0.05 كما نظن. وكلما ازداد الانحياز يصبح الاضطراب أكثر خطورة وعندما يكون

$B=\sigma$ فإن الاحتمال الكلي يصبح 0.17 أي أكثر من ثلاثة أمثال القيمة التي نظن ، ويختلف الذيلان في تأثرهما ، فمع الانحياز الموجب كما في المثال أعلاه، ينكمش احتمال تقدير بالنقصان يتجاوز 1.96σ انكماشاً سريعاً عن القيمة المفترضة 0.025 ليصبح مهملاً تقريباً عند $B=\sigma$. واحتمال التقدير بالزيادة المقابل يصعد بثبات. ويكون للخطأ الكلي الأهمية الأولى في معظم التطبيقات، ولكننا من حين لآخر، نهتم بصورة خاصة بالأخطاء في اتجاه معين.

وكقاعدة عامة ، نقول إن تأثير الانحياز في دقة تقدير ما يكون مهملاً إذا كان الانحياز أقل من عشر الانحراف المعياري لهذا التقدير وإذا كانت لدينا طريقة منحازة في التقدير وكان $\frac{B}{\sigma}<0.10$ حيث B هي القيمة المطلقة للتحيز ، فيمكننا الادعاء عندئذ بأن الانحياز لا يشكل عبأً أكبر لهذه الطريقة وحتى في حالة $\frac{B}{\sigma}=0.2$ يكون الاضطراب في احتمال الخطأ الكلي متواضعاً.

18-1 متوسط مربعات الخطأ:

كي نقارن تقديراً غير منحاز ، أو تقديرين منحازين بمقدارين مختلفين من الانحياز، يمكن استخدام قاعدة مفيدة هي متوسط مربعات خطأ التقدير (MSE) مقيساً بدءاً من قيمة المجتمع التي نريد تقديرها فنكتب:

$$
\begin{aligned}
MSE(\hat{\mu}) &= E(\hat{\mu}-\mu)^2 = E[(\hat{\mu}-m)+(m-\mu)]^2 \\
&= E(\hat{\mu}-m)^2 + 2\underbrace{(m-\mu)E(\hat{\mu}-m)}_{\text{معلوم لأن } E\hat{\mu}=m} + (m-\mu)^2 \\
&= E(\hat{\mu}-m)^2 + (m-\mu)^2 = [\text{تباين } \hat{\mu}] + [\text{الانحياز}]
\end{aligned}
$$

واستخدام متوسط مربعات الخطأ (MSE) كقاعدة لدقة مقدر يؤدي إلى افتراض تقديرين لهما الـ MSE نفسه كتقديرين متكافئين ، وهذا ليس صحيحا تماما وأن التوزيعين التكراريين لخطأين $(\hat{\mu}-\mu)$ مختلفين في حجميهما سوف لا يتطابقان من

أجل التقديرين إذا اختلفا في مقدار انحيازهما إلا أنه H. Madaw بين أنه إذا كـان $\frac{B}{\sigma}$ أقل من النصف تقريبا توزيعي التكرار يتطابقان تقريبا فيما يتعلـق بـالقيم المطلقـة لخطأين $(\hat{u}-\mu)$ من حجمين مختلفين ويوضح الجدول هذه النتيجة:

$\frac{B}{\sigma}$	$\sqrt{MSE}$	$1.96\sqrt{MSE}$	$2.576\sqrt{MSE}$
0	0.317	0.0500	0.0100
0.2	0.317	0.0499	0.0100
0.4	0.319	0.0495	0.0095
0.6	0.324	0.0479	0.0083

وحتى إذا كان $\frac{B}{\sigma}=0.6$ فإن التغيرات في الاحتمالات بالمقارنة مع الاحتمـال الموافق للحالة $\frac{B}{\sigma}=0$ هي تغيرات طفيفة. وبسبب صعوبة التأكد من عـدم وجـود انحياز أكيد في التقديرات سنتكلم عادة عن أحكام تقدير بدلا من دقة تقدير. فالدقـة تشير إلى حجم الانحراف عن المتوسط الصحيح μ. بينما يشير الإحكام إلى حجـم الانحراف عن المتوسط m الناتج عن تطبيق أسلوب المعاينة نفسه بصورة متكررة.

19-1 تمارين محلولة:

تمرين 1 :

يتكون مجتمع من خمسة أرقام 11 ,8 ,6 2,3 وتعد كل العينات الممكنة التي يكون حجمها 2 والمسحوبة على التوالي مع الإعادة من هذا المجتمع.

المطلوب :

a- عين متوسط المجتمع

b- عين الانحراف المعياري للمجتمع

C- عين متوسط توزيع المعاينة للأوساط

d- عين الانحراف المعياري لتوزيع المعاينة للأوساط أي الخطأ المعياري للأوساط

e- حل المسألة السابقة في حالة المعاينة بدون إعادة

الحل:

-a
$$\mu = \frac{2+3+6+8+11}{5} = \frac{30}{5} = 6$$

-b
$$\sigma^2 = \frac{(2-6)^2 + (3-6)^2 + (6-6)^2 + (8-6)^2 + (11-6)^2}{5} = 10.8$$

$$\sigma = \sqrt{10.8} = 3.29$$

c- لدينا هنا عدد العينات التي حجمها 2 والسحب مع الإعادة هو

$$n^r = 5^2 = 2$$

وهي:

(2,2), (2,3), (2,6), (2,8), (2,11)
(3,2), (3,3), (3,6), (3.8), (3,11)
(6,2), (6,3), (6,6), (6,8), (6,11)
(8,2), (8,3), (8,6), (8,8,), (8,11)
(11,2), (11,3), (11,6), (11,8), (11,11)

والأوساط المقابلة

2	2.5	4.0	5.0	6.5	
2.5	3.0	4.5	5.5	7.0	
4.0	4.5	6.0	7.0	8.5	(1)

5.0	5.5	7.0	8.0	9.5
6.5	7.0	8.5	9.5	11.0

والمتوسط الحسابي لتوزيع المعاينة للمتوسطات هو

$$\mu_{\bar{X}} = \frac{\sum_{i=m}^{25} \bar{X}_i}{25} = \frac{150}{25} = 6.0$$

وهذا يوضح أن $\mu_{\bar{X}} = \mu$.

d- التباين $\sigma_{\bar{X}}^{2}$ لتوزيع المعاينة للمتوسطات نحصل عليه بطرح المتوسط من كل رقم في (1) ، وتر بيع الناتج ثم جمع الحاصل والتقسيم على عدد العينات 25 فتكون:

$$\sigma_{\bar{X}}^2 = \frac{(2.-6)^2 + (2.5-6)^2 + (4-6)^2 + ...+ (11-6)^2}{25}$$

$$= \frac{135}{25} = 5.40 \Rightarrow \sigma_{\bar{X}} = \sqrt{5.40} = 2.32$$

وهذا يوضح حقيقة أن في المجتمعات المحدودة المتضمن المعاينة مع الإعادة :

$$\sigma_{\bar{X}}^{2} = \frac{\sigma^2}{n} = \frac{10.8}{2} = 5.4$$

وهي نتيجة مطابقة للقيمة إعادة:

e- في حالة المعاينة بدون إعادة:

لدينا من b, a : $\mu = 6; \sigma = 3.29$

ولإعادة C : لدينا عدد العينات التي حجمها 2 والسحب بدون إعادة

$$\binom{N}{n} = \binom{5}{2} = \frac{5!}{2!3!} = 10$$

(2,3), (2,6), (2,8), (2,11), (3,6), (3,8), (3,11)
(6,8), (6,11), (8,11)

والمتوسطات المقابلة هي :

2.5, 4.0, 5.0, 6.5, 4.5, 5.5, 7.0, 7, 8.5, 9.5

ومتوسط توزيع المعاينة للمتوسطات هو /

$$\mu_{\bar{X}} = \frac{2.5 + 4.0 + ... + 0.5}{10} = 6.0$$

وهذا يوضح حقيقة أن $\mu_{\bar{X}} = \mu$

ولإيجاد d: تباين توزيع المعاينة للمتوسطات هو

$$\sigma^2_{\bar{X}} = \frac{(2.5-6.0)^2 + (4.0-6.0)^2 + ...+ (9.5-6.0)^2}{10}$$

$$= 4.05 \Rightarrow \sigma_{\bar{X}} = 2.01$$

وهذا يوضح أن :

$$\sigma^2_{\bar{X}} = \frac{\sigma^2}{n}.(\frac{N-n}{N-1}) = \frac{10.8}{2}(\frac{5-2}{5-1}) = 4.05$$

وهذا كما حصلنا عليه أعلاه.

تمرين (2) :

لنفترض أن أوزان 3000طالب في جامعة يتوزع توزعا طبيعيا بمتوسط K.g 68 وبانحراف معياري3 kg سحبنا 80 عينة كل منها مكونة من 25 طالبا. عين المتوسط المتوقع والانحراف المعياري لتوزيع المعاينة للمتوسط إذا كانت المعاينة:

a- مع الإعادة

b- بدون إعادة

c- في كم من العينات نتوقع أن نجد المتوسط الحسابي

1- ما بين 66.8 و68.3 k.g

2- أقل من 66.4 k.g

الحل :

إن عدد العينات ذات الحجم 25 والتي يمكن الحصول عليها نظريا من مجموعة من 3000 طالب مع الإعادة 3000^{25} وبدون إعادة هو $\binom{3000}{25}$ وهو عدد أكبر من 80 وبهذا فإننا لم نحصل على توزيع المعاينة الحقيقي للمتوسطات ولكن نحصل على توزيع المعاينة التجريبي، وعلى الرغم من ذلك وبما أن عدد العينات كبير، فإننا نتوقع أن

يكون هناك اتفاق بين توزيعي المعاينة وبهذا فإن المتوسط المتوقع والانحراف المعيـــاري سيكونان قريبين من نظائرهما في التوزيع النظري ومنه فنحصل على:

-a $\mu_{\bar{X}} = \mu = 68 \quad k.g$

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{3}{\sqrt{25}} = 0.6 \quad k.g$$

-b $\mu_{\bar{X}} = \mu = 68 \quad k.g$

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} = \frac{3}{\sqrt{25}}.\sqrt{\frac{3000-25}{3000-1}} = 0.59$$

وهذا يختلف قليلا عن 0.6 k.g ويمكن بذلك عده لجميع الأغراض العملية مثل نظيرة في حالة المعاينة بإرجاع.

c- (1) : إن القيمة المعيارية لمتوسط العينة:

$$Z = \frac{X - \mu_{\bar{X}}}{\sigma_{\bar{X}}} = \frac{\bar{X} - 68}{0.6}$$

$$P(66.3 \leq \bar{X} \leq 68.3) = P(\frac{66.3-68}{0.6} \leq Z \leq \frac{68.3-68}{0.6})$$

$$= P(\quad -2.0 \leq Z \leq 0.5)$$

$$= P(Z \leq 0.5) - P(Z \leq \quad -2.0)$$

ومن جدول التوزيع الطبيعي المعياري نجد أنه:

$$= 0.6915 - 0.0228 = 0.6687$$

ويكون عدد العينات المتوقع هو: $|\Omega| = (80)(0.6687) = 53$

:(2)

$$P(\bar{X} < 66.4) = P(\frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} < \frac{66.4-68}{0.6})$$

$$= P(Z < -2.67) = 0.0638$$

وبهذا يكون العدد المتوقع للعينات

$$|\Omega| = (80)(0.0038) = 0.304 \# 0$$

تمرين (3):

لدينا 500 كرة حديدية متوسط وزنها 5.02 وانحرافها المعياري 0.30 سحبنا عينة عشوائية حجمها 100 كرة حديدية منها. عين احتمال أن يكون متوسط وزن الكرة

a- ما بين 4.96 و 5.00

b- أكثر من 5.10

الحل :

إن متوسط توزيع المعاينة للمتوسطات

$$\mu_{\bar{X}} = \mu = 5.02$$

والانحراف المعياري لتوزيع المعاينة للمتوسطات

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} = \frac{0.30}{\sqrt{100}} \cdot \sqrt{\frac{500-100}{500-1}} = 0.027$$

(a

$$P(4.96 \le \bar{X} \le 5.00) = P\left(\frac{4.96-5.02}{0.027} \le \frac{\bar{X}-\mu_{\bar{X}}}{\sigma_{\bar{X}}} \le \frac{5-5.02}{0.027}\right)$$

$$= P(-2.22 \le Z \le -0.74)$$

$$= P(Z \le -0.74) - P(Z \le -2.22)$$

$$= 0.2296 - 0.0132 = 0.2164$$

(b

$$P(\bar{X} > 5.10) = P\left(Z > \frac{5.10-5.02}{0.027}\right)$$

$$= P(Z > 2.96) = 1 - P(Z < 2.96)$$

$$= 1 - 0.9985 = 0.0015$$

c) إذا طلب إيجاد احتمال أن تكون أوزانها مجتمعة بين 496 و500 أو أكثر من .510

إن الوزن المجمع سوف يقع بين 496 و500 إذا كان متوسط وزن الـ 100 كرة يقع بين 4.96 و 5.00 وهذا الاحتمال حسبناه وهو 0.2164.

والوزن المجمع سوف يزيد على 510 إذا كان متوسط وزن الـ 100 كرة سوف يتجاوز 5.10 وهذا الاحتمال حسبناه وهو 0.0015.

تمرين (4):

ألقينا قطعة نقود 120 رمية عين احتمال الحصول على الصورة

a- ما بين 0.40 و 0.60 من المرميات

b- أكثر من $\frac{5}{8}$ من الرميات.

الحل:

نفترض أن الـ 120 رمية لقطة النقود كعينة من المجتمع غير المحدود المكون من جميع الرميات الممكنة للقطعة. وفي هذا المجتمع يكون احتمال الحصول على الصورة $P=\frac{1}{2}$ واحتمال الحصول على الكتابة هو $q=1-P=\frac{1}{2}$ وبالتالي :

a- المطلوب أن يكون عدد الصور في الـ 120 رمية بين 48 =(0.40)(120) و 72 = (0.60)(120) وبالتالي باستخدام التوزيع الطبيعي كتقريب لتوزيع ذي الحدين وافتراض X هو المتغير الدال على عدد الصور الحاصلة عندئذ

مستمر تقريب متقطع

$$P\ (48 \le X \le 72) \approx P(47.5 \le X \le 72.5)$$

$$= P\left(\frac{47.5-\mu_X}{\sigma_X} \le \frac{X-\mu_X}{\sigma_X} \le \frac{72.5-\mu_X}{\sigma_X}\right)$$

$$\mu_X = nP = (120)(\frac{1}{2}) = 60$$

$$\sigma_X = \sqrt{nPq} = \sqrt{(120)(\frac{1}{2})(\frac{1}{2})} = 5.48$$

$$P(48 \le X \le 72) = P\left(\frac{47.5-60}{5.48} \le Z \le \frac{72.5-60}{5.48}\right)$$

$$= P(-2.28 \le Z \le 2.28)$$

$$= P(Z \le 2.28) - P(Z \le -2.28)$$

$$= 0.9887 - 0.0113 = 0.97745$$

b- إن $\frac{5}{8} = 0.6250$ وعدد الصور يكون

$$(120)(0.6250) = 75$$

$$P(X \ge 75) = P(Z \ge \frac{74.5-60}{5.47})$$

$$= P(Z \ge 2.65) = 1 - P(Z < 2.65)$$

$$= 1 - 0.9960 = 0.0040$$

c- قام كل شخص من مجموعة مكون من 500 شخص يقذف قطعة نقد متوازنة 120 مرة. عين العدد المتوقع للأشخاص الذين يقرون أن رمياتهم أظهرت الصورة

1- ما بين 40% و60%.

2- $\frac{5}{8}$ أو أكثر

الجواب:

1- نفترض هنا أنه لدينا 500 عينة وحجم كل منها 120 مسحوبة من مجتمع غير محدود يمثل جميع الرميات الممكنة لقطعة النقد. وكما وجدنا من الطلب (a) وجدنا أن

الاحتمال الموافق لهذا الطلب وهو 0.9774 أي العدد المتوقع للأشخاص الذين يقـرون أن بين %40 و %60 من رمياهم أظهرت الصورة هو $489=(0.9774)(500)=|\Omega|$ شخص.

وتجدر الملاحظة هنا أن 11 = 489 -500 الباقين من الـ 500 شخص يقررون أن نسبة الصورة لا تقع بين 0.40 و0.60 مثل هؤلاء الأشخاص قد يدعون أن قطعـات النقد التي ألقوها غير متوازنة على الرغم من أنها ليست كذلك ، وهذا النـوع مـن الخطأ هو المخاطرة التي تظهر كلما تعاملنا مع الاحتمالات.

2- وبالمسوغات نفسها كما في الطلب السابق نستنتج أن

2= (0.0040) (500) شخص سوف يقررون أن $\frac{5}{8}$ أو أكثر من رميـاهم ينتـج عنها ظهور الصورة:

جدول الأرقام العشوائية

51772	74640	23315	29049	2898	93521
04167	108638	76109	19605	5604	06563
24031	23490	45931	60172	52083	11645
30587	21333	75790	45402	31414	67076
64934	51135	30277	94623	85418	68829
57683	30277	94623	85418	68829	06652
56301	09448	21631	91157	77331	60710
73547	76552	50020	24819	1853	52290
42457	95652	25496	50532	07136	40876
79971	54195	25708	51817	36732	72484
27989	64728	10766	08299	58100	90012
28184	99230	85077	73370	36189	46711
00398	25314	09998	36346	59441	55650
11154	32542	36749	97000	21114	54431
05128	23001	49001	81044	72112	35200
20261	30101	51887	91442	82031	69838
00304	28332	70111	68323	42009	53107
71410	89560	90232	04918	12801	19324
31587	74790	00446	28111	15400	34981
03891	00924	86557	83203	15055	16416

30727	79148	12668	38181	29927	96393
44008	88362	60870	51277	23769	18265
31332	87586	78984	97563	38388	31005
34917	65700	67028	04678	90141	16240
30834	92914	50142	95816	07539	03414
53692	90803	29461	64986	09681	85632
48454	50643	17751	37028	10708	04158
81310	12987	01812	43061	61339	87439
23621	34676	73224	21070	30832	15017
40933	96456	45448	05489	04792	50010
69045	89238	76513	98792	35341	04752
12281	68659	00004	25115	31158	77251
38427	77713	38219	73300	72185	46552
23191	56811	3474	44111	81090	07150
07418	85029	29423	36322	39329	01862
04760	03544	00752	88285	48848	40291
46910	40330	8101	00739	17602	7244
55144	92158	99218	12800	06587	10423
30036	01272	10020	00551	45339	88112
41107	02223	76823	93000	12991	26435
13489	33690	63985	98467	72680	24210
17591	24411	52066	12753	98058	42614
98972	86342	74431	50601	11269	99915
89758	27038	41318	87704	93245	32888
00646	05217	34018	75623	83358	55412
41514	10904	05217	26521	57857	47279
94524	15305	016816	36491	97531	86420

20-1 : تمارين غير محلولة:

تمرين (1):

1- عدد ميزات نظرية العينات، ثم عدد مجالات تطبيق نظرية العينات.

2- عرف كلا من المجتمع والإطار ووحدة المعاينة، ثم عدد الشروط الأساسية للمعاينة.

3- ما هي أشكال سحب العينات وعدد طرائق سحب العينة.

4- اشرح طريقة السحب بوساطة جداول الأرقام العشوائية واسحب منها رقما ذا ثلاث مراتب.

5- اشرح طريقة السحب بوساطة تطبيق طريقة مونتن – كارلو وارسم مخطط العمليات الحسابية لسحب العينات بوساطة تطبيق طريقة مونتي كارلو.

6- عدد الخطوات الأساسية لتصميم العينة

7- عدد معايير جودة التقدير وعرف كلا منها.

8- ما هو عدد العينات الممكن سحبها من مجتمع N وبحجم n.

وما هو الشكل العام لتوزيع متوسطات العينات.

9- احسب احتمال وجود عنصر ما في العينة ذات الحجم n من مجتمع N.

10- عرف تباين متوسطات العينات.

11- عرف كلا من الخطأ المطلق والدقة وأوجد العلاقة بينهما.

12- عرف نسبة وجود خاصة معينة من العينة والمجتمع.

13- احسب احتمال ظهور عنصر ما ثلاث مرات في عينة ذات حجم 10 واحدات.

14- احسب احتمال ظهور عنصر ما ثلاث مرات على الأكثر في عينة ذات حجم 10 واحدات.

15- احسب احتمال ظهور عنصر ما ثلاث مرات على الأقل في عينة ذات حجم 10 واحدات.

تمرين (2):

ستأخذ عينة من قائمة من الأسماء المسجلة على بطاقات (اسم في كل بطاقة) مرقمة على التسلسل في إضبارة. ولكل اسم الفرصة نفسها في أن يسحب في العينة.

ما هي المشكلات التي تنشأ في الحالات العامة التالية؟

a- بعض الأسماء لا تنتمي إلى المجتمع الهدف، علما بأن هذه الحقيقة لا يمكن التأكد منها إلا بعد سحب الاسم.

b- بعض الأسماء تظهر على أكثر من بطاقة ، وجميع البطاقات التي تحمل الاسم نفسه لها أرقام متسلسلة وبالتالي تظهر بعضها إلى جانب بعض في الإضبارة.

c- بعض الأسماء تظهر على أكثر من بطاقة والبطاقات التي تحمل الاسم نفسه مبعثره ضمن الإضبارة.

تمرين (3):

تشكل مسألة إيجاد وإطار كامل يسمح بسحب العينة عقبة في الغالب. ما هي أنواع الإطارات التي يمكن تجريبها في المسوح التالية؟ وهل لهذه الإطارات أي نقاط ضعف خطيرة؟

a- مسح للمخازن التي تبيع حقائب سفر في مدينة كبيرة

b- مسح لأنواع المواد التي تركها أصحابها في قطارات الأنفاق أو الحافلات العامة.

c- مسح للأشخاص الذين لدغتهم الثعابين في العام الماضي.

d- مسح لتقدير عدد الساعات الأسبوعية التي تقضيها أفراد أسرة في مشاهدة التلفزيون.

تمرين (4): دليل مدينة عمره أربع سنوات ويتضمن العناوين مرتبة على طول كل شارع ، كما يعطي أسماء الأشخاص الذين يعيشون في كل عنوان. يراد أجراء مسح للأشخاص في المدينة ، يجري بطريقة المقابلة ، ما هو نواقص هذا الإطار ؟ هل يمكن معالجة هذه النواقص من قبل العداد خلال قيامهم بعملهم الميداني! عند استخدامك للدليل ، هل تسحب قائمة من العناوين (أمكنة السكن) أم قائمة من الأشخاص؟

تمرين (5):

عند تقدير القيمة الفعلية للبنود الصغيرة في مستودعات شركة كبيرة بطريقة العينة، سجلنا القيمة الفعلية إلى القيمة الدفترية لكل البنود في العينة. ووجدنا أن نسبة القيمة الفعلية إلى القيمة المسجلة في العينة كلها كانت (1.021) وهذا التقدير يتوزع طبيعا بانحراف معياري 0.0082. إذا كانت القيمة الدفترية للمخزون المراد تقدير قيمته الفعلية هي 80000 دولار. فاحسب 95% حدود ثقة للقيمة الفعلية.

تمرين (6):

كثيرا ما نتعامل مع بيان إحصائي كعينة ، مع أنها تبدو للوهلة الأولى وكأنها حصر شامل. ويجد صاحب موقف السيارات ضآلة العمل في أيام الأحد صباحا. وبعد ستة وعشرين يوما (أحد) من العمل كان متوسط ما تسلمه صباح الأحد هو 10 دولار بالضبط. ويتقاضى الحارس 7 دولارات كل أحد. ويرحب المالك بترك الموقف مفتوحا للسيارات صباح الأحد إذا كان توقع ربحه في المستقبل يبلغ الـ 5 دولارات كل صباح أحد . ما معامل الثقة الاحتمالية بأن معدل ربحه على المدى الطويل سيكون 5 دولارات على الأقل؟

وما هي الفرضيات التي يجب وضعها للإجابة عن السؤال؟

تمرين (7):

يتكون مجتمع من أربعة أرقام 15 ,11 ,7 ,3 لنعد كل العينات الممكنة ذات الحجم n=2 والتي يمكن سحبها مع الإعادة من هذا المجتمع. والمطلوب

1- عين متوسط المجتمع وانحرافه المعياري

2- عين متوسط توزيع المعاينة للمتوسطات وانحراف المعياري

3- حل المسألة في حالة السحب بدون إعادة.

تمرين (8):

يتوزع وزن كرة حديدية من مجتمع من 1500 كرة طبيعيا بتوقع 22.40 وانحراف معياري 0.048 سحبت 300 عينة حجم كل منها 36 من هذا المجتمع.

1- عين المتوسط المتوقع والانحراف المعياري لتوزيع المعاينة للمتوسطات إذا كانت المعاينة.

a- مع الإعادة

b- بدون إعادة.

2- كم من العينات العشوائية متوسطاتها تقع بين 22.39 و 22.41 نيوتن.

3- أكبر من 22.42.

تمرين (9):

إذا كان متوسط وزن طرود مرسلة إلى أحد المتاجر هو 300 بانحراف معياري 50. اختير 25 طردا بصورة عشوائية ووضعت في مصعد لرفعها. عين احتمال أن يكون وزن الطرود سوف يتجاوز حدود الأمان المحددة للصعود والمقرره بـ 8200؟

تمرين (10):

عين احتمال أن يكون بين 400 طفل سوف يولدون:

a- أقل من 40% سيكونون ذكورا

b- بين 43% و57% سيكونون إناثا

c- أكبر من 54% سيكونون ذكورا.

تمرين (11):

صندوق يحوي 80 كرة متماثلة فيها 60% من اللون الأحمر و40% من اللون الأبيض.

سحبنا 50 عينة كل منها مؤلف من 20 كرة مع الإرجاع من هذا الصندوق. كم من العينات نتوقع أن تتكون :

a- من عدد متساو من الكرات الحمراء والبيضاء

b- 15 حمراء و5 بيضاء

c- 10 أو أكثر من الكرات حمراء.

d- 7 بيضاء و13 حمراء.

تمرين (12):

أرسل مصنع 1000 طرد يتكون كل منها من 100 مصبـــــاح كـــهربائي، وإذا كانت 0.05 من المصابيح تالفه، والتوزيع طبيعي ، عين عدد الطرود التي نتوقـــــع أن يكون فيها.

a- أقل من 90 مصباحا صالح للعمل

b- 98 أو أكثر صالح للعمل

تمرين (13):

اكتب العلاقة الدالة على توليد الأعداد شبه العشوائية الصحيحة ثم طبق هـــــذه العلاقة في توليد عينة عشوائية مؤلفة من ثلاثة عناصر من مجتمع من 6000 عنصر مـــع العلم أن $n_0 = 1111$ و $\varepsilon = 1112$

الجواب :

$$n_1 = 2666 \quad ; n_2 = 274 \qquad n_3 = 649$$
$$n_4 = 9828 \quad ; n_5 = 313$$

تمرين (14) :

اكتب علاقة نيمان .الدالة على توليد الأعداد شبه العشوائية واشرح محتويات هذه العلاقة وطبقها في تشكيل عينة عشوائية حجمها 7 من مجتمعا مؤلفا من 500 عنصـــــر علما بأن $V_0 = 0.1111$

الجواب :

118; 245; 486; 123; 7; 8; 12

تمرين (15):

عدد طرائق سحب العينات العشوائية واشرح طريقة مونتي كارلو ثم اشرح علاقة ينمان في توليد الأعداد شبه العشوائية.

تمرين (16):

مجتمع إحصائي مؤلف من العناصر التالية:

1,2,3,4,5 ولندرس كل العينات من الحجم 2 والممكن ســـــحبها مـــن المجتمـــع المفروض والمطلوب:

1- عين متوسط المجتمع وانحرافه المعياري

2- عين توقع متوسط المعاينة وانحرافه المعياري وماذا نستنتج وذلك في الحـــالات التالية:

a- سحب مع الإعادة بدون تمييز

b- سحب بدون إعادة.

تمرين (17):

لنفترض أن أطوال 2000 طالب في كلية العلوم تتوزع توزعا طبيعيا بمتوسط 170 c.m ، وانحراف معياري 10 c.m . إذا سحبت 100 عينة كل منها مؤلف مـــن 100 طالب. عين توقع متوسط المعاينة وانحرافه المعياري وفي كم من العينات نتوقع أن نجـد متوسط الطول.

1- محصورا بين 160 و170 C.m

2- أقل من 165 C.m

3- أكثر من 172 C.m

تمرين (18):

ليكن لدينا مجتمع مؤلف من 3000 عنصر ونريد سحب عينة منـــه حجمـــها 3 عناصر. وذلك حسب خوارزمية توليد الأعداد شبه العشوائية الصحيحة ، علما بأن

$\varepsilon=1100 \quad ; n_0=100$

الجواب :

$n_1=2410 \; ; \; n_2=1196 ; \qquad n_3=8362 ; n_4=7040$

$n_5=1070$

الفصل الثاني
المعاينة العشوائية البسيطة

2-1- مقدمة:

المعاينة العشوائية البسيطة هي طريقة لاختيار n وحده من بين N بحيث يكون لكل من العيّنات الـ $\binom{N}{n}$ الممكنة الفرصة نفسها في أن تكون هي العينة المسحوبة وبالتالي فهي التصميم الذي يتساوى فيه احتمال انتقاء أي من العينات ذات الحجم n الممكنة التشكيل من مجتمع مؤلف من N عنصراً. وتطبيق هذه المعانية في المجتمعات المتجانسة من حيث قيّم الخاصة المدروسة. وعملياً تسحب العينة العشوائية البسيطة وحده فوحده، ونرّقم الوحدات في المجتمع من 1 إلى N ثم نسحب سلسلة من الأعداد العشوائية بين 1 وN، أما بوساطة جداول الأرقام العشوائية أو بوساطة برنامج على الحاسوب يستنتج مثل هذه الجداول. وعند كل سحب يجب أن تعطي الطريقة المستخدمة فرصة الاختيار نفسها لأي عنصر من المجتمع لم يجر سحبه بعد.

فمن أجل مجموعة من n وحدة محددة ، يكون احتمال اختيار واحدة من هذه الوحدات في السحب الأول هو $\frac{n}{N}$ وفي السحب الثاني $\frac{n-1}{N-1}$ وهكذا ....
وبالتالي فاحتمال اختيار وحدات العينة الـ n خلال n سحبا يكون:

$$\frac{n}{N}\cdot\frac{(n-1)}{(N-1)}\cdot\frac{(n-2)}{(N-2)}\cdots\cdots\frac{1}{N-n+1}=\frac{n!(N-n)!}{N!}=\frac{1}{\binom{N}{n}} \qquad (1)$$

وبما أننا نخرج من المجتمع الرقم الذي يجري سحبه وذلك في كل عملية من عمليات السحب اللاحقة فتدعى هذه الطريقة بالمعاينة بدون إعادة.

أما المعاينة مع الإعادة فتكون كما يلي: عند كل سحب يُعطى كل عدد من الأعداد الـ N في المجتمع الفرصة نفسها في أن يكون هو العدد المسحوب، وذلك

بصرف النظر عن تكرار سحب أي عدد. والعلاقات الخاصة بالتباينات وتقدير تباينات التقديرات التي نحسبها من العينة، غالباً ما تكون في المعاينة مع الإعادة أبسط منها في المعانية بدون إعادة. ولهذا السبب نستخدم أحياناً المعاينة مع الإعادة في خطط المعاينة الأكثر تقيداً. وهذه التقديرات نحصل عليها بدراسة واحدات العينة المسحوبة ثم حساب الوسطاء المقابلة لها في مجتمع العينة، وللتغلب على مخاطر الانتقال يجب أن ننطلق من المعايير الأساسية وهي عدم التحيّز والتماسك والفعالية للتقدير.

2-2: تعاريف ورموز:

نرمز لواحدات المجتمع الـ N بـ $y_1, y_2, \ldots\ldots, y_N$ أو y_i : $i = 1, 2, \ldots\ldots, N$

ونشير إلى الصفات المميّزة لمجتمع بأحرف كبيرة وللصفات المميّزة لعينة بأحرف صغيرة ، ونرمز لـ: $\sum_{i=1}^{n} y_i = y_1 + y_2 + \ldots + y_n$ لمجموع واحدات العينة

$Y = \sum_{i=1}^{N} y_i = y_1 + y_2 + \ldots + y_N$ لمجموع واحدات المجتمع

$\bar{y} = \frac{y_1 + y_2 + \ldots + y_n}{n} = \frac{\sum_{i=1}^{n} y_i}{n}$ لمتوسط العينة

$\bar{Y} = \frac{y_1 + y_2 + \ldots + y_N}{N} = \frac{\sum_{i=1}^{N} y_i}{N}$ لمتوسط المجتمع

والاهتمام بتركز في معظم الأحيان على أربع صفات مميّزة للمجتمع:

1- المتوسط $\bar{Y}$ (مثلاً معدل عدد الأطفال في المدرسة الواحدة).

2- المجموع Y (مثلاً مجموع عدد دونمات القمح في المنطقة).

3- نسبة مجموعين أو متوسطين $R = \frac{Y}{X} = \frac{\bar{Y}}{\bar{X}}$

(مثلاً: نسبة الممتلكات المنقولة إلى الممتلكات الكلّية في مجموعة من الأسر).

4- نسبة الواحدات التي تقع ضمن صنف معيّن (مثلاً نسبة الأشـــخاص الذيـــن يمتلكون أسناناً صناعية).

وسوف نستخدم الرمز A للدلالة على تقدير حصلنا عليه من العينـــة لإحـــدى الصفات المميّزة للمجتمع وسوف نأخذ أبسط أنواع التقديرات هنا.

المقدّر $\hat{\bar{Y}}=\bar{y}$ لـ $\bar{Y}$ أي متوسط العينة مقدر لمتوسط المجتمع

والمقدّر $\hat{Y}=N\bar{y}$ لـ Y (للمجموع الكلي للمجتمع)

والمقدّر $\hat{R}=\frac{\bar{y}}{\bar{X}}=.\frac{\sum_{1}^{n}y_i}{\sum_{1}^{n}n_i}$ لـ R (للنسبة في المجتمع)

ومن أجل $\bar{Y}=N\bar{y}=\frac{N}{n}\sum_{i=1}^{n}y_i$ فإن $\frac{N}{n}$ يدعى بعامل التوسع أو عام النهوض أو عامل التضخم . وعكسه $\frac{n}{N}$ يدعى بكسر المعاينة ونرمز له بـــ f .

2-3: خواص التقديرات:

تعتمد دقة أي تقدير نقوم به من العينة على الطريقة التي تحسب فيها هذا التقديـــر من البيان الإحصائي للعينة، وعلى خطة المعاينة ونكتب أحيانا للاختزال "دقه متوسط العينة أو دقة المعاينة العشوائية البسيطة".

وهنا سنقول إن طريقة المعاينة متسقه إذا أصبح التقدير مساويا تماما للقيمة المقدرة من المجتمع وذلك عندما تصبح N=n أي عندما تتألف العينة من المجتمـــــع بكاملـــه. ويكون $\bar{y}$, $N\bar{y}$ تقديرين متسقين لمتوسط المجتمع ومجموع المجتمع على الترتيب.

2 .3 .1 مبرهنة (1):

متوسط العينة $\bar{y}$ هو تقدير غير منحاز لمتوسط المجتمع $\bar{Y}$.

الاثبات: نعلم أن عدد العينات الممكن تشكليها في حالة السحب بدون إعـــادة يساوي $\binom{N}{n}$. واحتمال انتقـــاء أي مـــن هـــذه العينـــات متســـاو ويســـاوي $P_n=\frac{1}{\binom{N}{n}}$ وبالتالي:

$$E(\bar{y})=\frac{\sum_{K=1}^{\binom{N}{n}}\bar{y}_K}{\binom{N}{n}}=\frac{\sum_{K=1}^{\binom{N}{n}}(y_1+y_2+....+y_n)_K}{n\left[\frac{N!}{n!(N-n)!}\right]} \qquad (2)$$

ولحساب المجموع نبحث عن عدد العيّنات التي تظهر فيها أي قيمة معينـــة y_i .
وحيث إن هناك (N-1) وحدة أخرى متاحة لاستكمال العينة وكذلك (n-1) خانـــة أخرى من العينة ينبغي شغلها فإن عدد العينات التي تحوي y_i هو $(i=1,2,......,N)$

$$\binom{N-1}{n-1}=\frac{(N-1)!}{(n-1)!(N-n)!} \qquad (3)$$

ومنه :

$$\sum_{K=1}^{\binom{N}{n}}(y_1+y_2+......+y_n)_K=\frac{(N-1)!}{(n-1)!(N-n)!}(y_1+y_2+.....+y_N)$$

$$E(\bar{y})=\frac{(N-1)!}{(n-1)!(N-n)!}.\frac{n!(N-n)!}{nN!}.(y_1+y_2+....+y_N)$$

$$=\frac{y_1+y_2+....+y_N}{N}=\bar{Y} \qquad (4)$$

وهنا يمكن أن نفسر عملية الحصول على عدد العينات التي تظهر فيها أي قيمـــة معينة y_i حيث إننا إذا حددنا عنصرا ما من المجتمع وليكن y_i فإنه يمكننا أن نحســب متكرارات هذا العنصر فيما إذا عرفنا عدد العينات التي ينتمي إليها هـــذا العنصـــر، ولذلك إذا افترضنا أن y_i ينتمي إلى العينة المراد سحبها (كعنصر أول مثلا) وأردنـــا أن نكمل عملية السحب حتى نشكل عينة ذات حجم n فإن عناصر المجتمع المتبقيـــة يصبح عددها (N-1) ويجب أن نسحب (n-1) عنصرا وهذا يعني أنه يجب أن نشـــكل عينات ذات حجم (n-1) من مجتمع من (N-1) ومنه سيكون عدد العينات التي تحـوي العنصر y_i مساويا عدد العينات التي يمكن تشكليها من (N-1) عنصرا وذات حجـــم

(n-1) أي أن عدد هذه العينات يساوي $\binom{N-1}{n-1}$. وبالتالي كل عنصر y_i من عناصر المجتمع يتكرر $\binom{N-1}{n-1}$ مرة تحت إشارة المجموع Σ.

نتيجة: إن $\hat{Y}=N\bar{y}$ هو تقدير غير منحاز لمجموع المجتمع Y

$$\left(Y=N\bar{Y}\Rightarrow\hat{Y}=N\hat{\bar{Y}}=N\bar{y}\ \text{لأن}\ \right) \qquad (5)$$

2-3-2: تباينات التقديرات:

نعرّف عادة تباين y_i $\quad i=1,.....,N$ في مجتمع منته بالعلاقة

$$\sigma^2=\frac{\sum_{i=1}^{N}(y_i-\bar{Y})^2}{N} \qquad (6)$$

وسنصطلح بتقديم النتائج بدلالة عبارة مختلفة قليلا يكون فيها المقام (N-1) بدلالة N، فنأخذ

$$S^2=\frac{\sum_{i=1}^{N}(y_i-\bar{Y})^2}{N-1} \qquad (7)$$

والعلاقة (7) تستعمل من قبل مستخدي تحليل التباين، وفائدتها هي أن معظم النتائج تأخذ شكلا أبسط بقليل، ويتقى النتائج نفسها عند استخدام أي من المصطلحين شريطة الثبات على استخدام الرمز نفسه.

ولندرس الآن تباين $\bar{y}$ أي $E(\bar{y}-\bar{Y})^2$ محسوبا من كل الـ $\binom{N}{n}$ من العينات الممكنة.

2-3-3: مبرهنة (2):

تباين المتوسط $\bar{y}$ لعينة عشوائية بسيطة هو

$$V(\bar{y})=E(\bar{y}-\bar{Y})^2=\frac{S^2}{n}.\left(\frac{N-n}{N}\right)=\frac{S^2}{n}(1-f) \qquad (8)$$

حيث $f=\frac{n}{N}$ يمثل كسر المعاينة.

الإثبات : لنأخذ العلاقة :

$$n(\bar{y}-\bar{Y})=(y_1-\bar{Y})+(y_2-\bar{Y})+.....+(y_n-\bar{Y}) \qquad (9)$$

وبسبب التناظر نجد:

$$E(y_1+y_2+....+y_n)=\frac{n}{N}(y_1+y_2+...+y_N)$$

حيث العبارة يساراً فيها n حداً ويميناً فيها N حداً ولأن :

$$E(\bar{y})=\bar{Y} \Rightarrow E(\sum_{i=1}^{n} y_i)=n\bar{Y}=n\frac{(\sum_{i=1}^{N} y_i)}{N}$$

فنجد عندئذ:

$$E\left[(y_1-\bar{Y})^2+....+(y_n-\bar{Y})^2\right]=\frac{n}{N}\left[(y_1-\bar{Y})^2+...+(y_N-\bar{Y})^2 \qquad -10\right.$$

ونجد أيضا :

$$E\left[(y_1-\bar{Y})(y_2-\bar{Y})+(y_1-\bar{Y})(y_3-\bar{Y})+......+(y_{n-1}-\bar{Y})(y_n-\bar{Y}\right] \qquad -11$$

$$=\frac{n(n-1)}{N(N-1)}\left[(y_1-\bar{Y})(y_2-\bar{Y})+(y_1-\bar{Y})(y_3-\bar{Y})+...+(y_{N-1}-\bar{Y})(y_N-\bar{Y})\right]$$

وفي المعادلة (11) تمتد المجاميع للجداءات فوق جميع الأزواج من الوحـــدات في العينة والمجتمع على الترتيب ويحوي المجموع الذي على اليسار $\frac{n(n-1)}{2}$ من الحـــدود والذي على اليمين $\frac{N(N-1)}{2}$ حدا.

لنربع الآن العلاقة (9) ولنسحب معدلها فوق جميع العينات العشوائية البســيطة ، وباستخدام (10) و (11) نحصل على :

$$n^2E(\bar{y}-\bar{Y})^2=\frac{n}{N}\left\{\left[(y_1-\bar{Y})^2+....+(y_N-\bar{Y})^2\right]\right.$$

$$\left.+\frac{2(n-1)}{N-1}\left[(y_1-\bar{Y})(y_2-\bar{Y})+.....+(y_{N-1}-\bar{Y})(y_N-\bar{Y})\right]\right\}$$

وبإتمام المربعات بالنسبة لكل حد جدائي يمكن أن نكتب

$$n^2 E(\bar{y}-\bar{Y})^2 = \frac{n}{N}\left\{ \left[(1-\frac{n-1}{N-1})[(y_1-\bar{Y})^2 + ... + (y_N-\bar{Y})^2] \right] + \right.$$

$$\left. \frac{(n-1)}{N-1}\left[(y_1-\bar{Y}) + ... + (y_N-\bar{Y})\right]^2 \right\}$$

ينعدم الحد الثاني داخل القوس المربع باعتبار أن مجمـــــوع y_i يســـاوي $N\bar{y}$.
وبقسمة الطرفين في العلاقة الأخيرة على (n^2) نجد

$$V(\bar{y}) = E(\bar{y}-\bar{Y})^2 = \frac{n}{n^2 N}.\left[\frac{(N-1)-(n-1)}{N-1}\right]\sum_{i=1}^{N}(y_i-\bar{Y})^2$$

$$= \frac{(N-n)}{nN(N-1)}.\sum_{i=1}^{N}(y_i-\bar{Y})^2 = \frac{S^2}{n}.\frac{N-n}{N}$$

$$= \frac{S^2}{n}.(1-f)$$

4-3-2 : نتائج:

نتيجة (1): الخطأ المعياري لـــ $\bar{y}$ وهو

$$\sigma_{\bar{y}} = \frac{S}{\sqrt{n}}.\sqrt{\frac{N-n}{N}} = \frac{S}{\sqrt{n}}.\sqrt{1-f} \qquad (12)$$

نتيجة (2) : تباين $\hat{Y} = N\bar{y}$ كتقدير لمجموع المجتمع Y هو :

$$V(\hat{Y}) = E(\hat{Y}-Y)^2 = V(N\bar{y}) = N^2 V(\bar{y})$$

$$= (\frac{N^2.S^2}{n}).(\frac{N-n}{N}) = \frac{N^2 S^2}{n}(1-f)$$

$$= \frac{N(N-n)}{n}.S^2 \qquad (13)$$

نتيجة (3): الخطأ المعياري لـــ $\hat{Y}$ هو

$$\sigma_{\hat{y}} = \frac{NS}{\sqrt{n}}.\sqrt{\frac{N-n}{N}} = \frac{NS}{\sqrt{n}}.\sqrt{1-f} \qquad (14)$$

2-3-5: التصحيح في حالة مجتمع منته:

نعلم أن تباين متوسط عينة عشوائية حجمها n من مجتمع لانهائي هو $\frac{\sigma^2}{n}$ والتغير الوحيد في هذه النتيجة عندما يكون المجتمع منتهيا يتحتم علينا إدخال العامل الإضـــافي $\frac{N-n}{N}$.

ويدعى العاملان $\frac{N-n}{N}$ ، $\sqrt{\frac{N-n}{N}}$ في حالة التباين والخطأ المعيـــاري على الترتيب بعاملي التصحيح لمجتمع منته وبكتاب بمقام (N-1) بدلا من N من قبـــل الكتاب الذين يقدمون النتائج بدلالة σ.

وهذان العاملان قريبان من الواحد شريطة أن يبقى كسر المعاينة $\frac{n}{N}$ منخفضـــا ، وهكذا فإنه لا يوجد لحجم المجتمع أي تأثير مباشر في الخطأ المعياري لمتوسط العينـــة. وعلى سبيل المثل إذا كانت S نفسها في المجتمعين ، فإن عينة حجمها 500 من مجتمع مؤلف من 200000 تعطي تقديرا لمتوسط المجتمع ، دقته تقريبا هي الدقة نفسها لتقدير عينة حجمها 500 من مجتمع من 10000. وفي التطبيقات يمكن إهمال معامل التصحيح لمجتمع منته حينما لا يتجاوز كسر المعاينة 5% ، ولغايات عديدة، حتى إذا كان عاليـــا حتى 10%. وتأثير تجاهل التصحيح هو المبالغة في تقدير الخطأ المعياري للتقدير $\bar{y}$.

2-3-6: مبرهنة (3):

إذا كان y_i ، x_i زوجا من المتغيرات معرفا في وحـــدة في المجتمـــع و $\bar{y}$ ، $\bar{x}$ يمثلان المتوسطين الموافقين لعينة عشوائية بسيطة حجمها n.

عندئذ التغاير:

$$E(\bar{y}-\bar{Y})(\bar{x}-\bar{X})=\frac{N-n}{nN}\cdot\frac{1}{N-1}\sum_{i=1}^{N}(y_i-\bar{Y})(x_i-\bar{X}) \quad (15)$$

وتختزل هذه المبرهنة إلى المبرهنة (2) إذا بقيت قيم المتغيرين y_i ، x_i نفســـها في كل وحده.

الإثبات: نطبق المبرهنة (2) على المتغير $u_i = y_i + x_i$ فمتوسط المجتمع لـ u_i هو $\overline{U} = \overline{Y} + \overline{X}$ وبالتالي:

$$E(\overline{u} - \overline{U})^2 = \frac{N-n}{nN} \cdot \frac{1}{N-1} \sum_{i=1}^{N} (u_i - \overline{U})^2$$

أي أن :

$$E\left[(\overline{y} - \overline{Y}) + (\overline{x} - \overline{X})\right]^2 = \frac{N-n}{nN} \cdot \frac{1}{N-1} \sum_{i=1}^{N} \left[(y_i - \overline{Y}) + (x_i - \overline{X})\right]^2 \qquad (16)$$

وبنشر المربعين في كلا الطرفين واستنادا للمبرهنة (2): نجد أن:

$$E(\overline{y} - \overline{Y})^2 = \frac{N-n}{nN} \cdot \frac{1}{N-1} \sum_{i=1}^{N} (y_i - \overline{Y})^2$$

$$E(\overline{x} - \overline{X})^2 = \frac{N-n}{nN} \cdot \frac{1}{N-1} \sum_{i=1}^{N} (n_i - \overline{X})^2$$

وهذه تحدد من الطرفين بعد نشر المربعين وتبقى الحدود الجدائية.

$$E(\overline{y} - \overline{Y})(\overline{x} - \overline{X})^2 = \frac{N-n}{nN} \cdot \frac{1}{N-1} \sum_{i=1}^{N} (y_i - \overline{Y})(x_i - \overline{X})$$

وهو المطلوب.

4-2: تقدير الخطأ المعياري من العينة:

نستخدم عادة، العلاقتين الموافقتين للخطأ المعياري لكل من تقدير متوسط مجتمع وتقدير مجموع مجتمع لغايات ثلاث وهي:

1- مقارنة الدقة الناتجة عن المعاينة العشوائية البسيطة مع تلك الناتجة عن طرائق أخرى في المعاينة.

2- تقدير حجم العينة الذي نحتاجه في مسح نقوم بتخطيطه.

3- تقدير الدقة الفعلية التي بلغناها في مسح تم إنجازه. والعلاقات تحوي S^2 (تباين المجتمع) حيث يمكن تقديره من البيانات الإحصائية في العينة.

1-4-2: مبرهنه (4): في حالة عينة عشوائية بسيطة يكون $s^2 = \frac{\sum_1^n (y_i - \bar{y})^2}{n-1}$

تقديراً غير منحاز لـــ $S^2 = \frac{\sum_{i=1}^N (y_i - \bar{Y})^2}{N-1}$

الإثبات: لدينا

$$s^2 = \frac{1}{n-1}\sum_{i=1}^n \left[(y_i - \bar{Y}) - (\bar{y} - \bar{Y})\right]^2 \qquad (17)$$

ومنه بفك التربيع والاختزال:

$$s^2 = \frac{1}{n-1}\left[\sum_{i=1}^n (y_i - \bar{Y})^2 - n(\bar{y} - \bar{Y})^2\right] \qquad (18)$$

والآن لنأخذ المعدّل فوق جميع العينات العشوائية البســـيطة الـــتي حجمـــها n، وبالاستناد إلى حجة التناظر ذاتها والمستخدمة في المبرهنة (2) نجد:

$$E\left[\sum_{i=1}^n (y_i - \bar{Y})^2\right] = \frac{n}{N}\sum_{i=1}^N (y_i - \bar{Y})^2 = \frac{n(N-1)}{N}S^2$$

وذلك من تعريف S^2 وبالإضافة لذلك نجد من المبرهنة (2).

$$E\left[n(\bar{y} - \bar{Y})^2\right] = \frac{N-n}{N}S^2$$

$$E(s^2) = \frac{1}{n-1}\left[\frac{n(N-1)}{N}S^2 \; \frac{N-n}{N}S^2\right]$$

$$= \frac{S^2}{N(n-1)}\left[nN - n - N + n\right] = S^2 \qquad (19)$$

2-4-2: نتيجة : التقديران غير المنحازين لتبايني $\bar{Y}$, $\hat{Y}=N\bar{Y}$ هما

$$V(\bar{Y})=S_{\bar{Y}}^2=\frac{s^2}{n}(\frac{N-n}{N})=\frac{s^2}{n}(1-f) \quad (20)$$

$$V(\hat{Y})=S_{\hat{Y}}^2=\frac{N^2 s^2}{n}(\frac{N-n}{N})=\frac{N^2 s^2}{n}(1-f) \quad (21)$$

ونأخذ في حالة الأخطاء المعيارية

$$S\bar{y}=\frac{s}{\sqrt{n}}.\sqrt{1-f}$$

$$S\hat{y}=\frac{Ns}{\sqrt{n}}.\sqrt{1-f} \quad (22)$$

وهذان التقديران منحازان بصورة طفيفة، وفي معظم التطبيقات العملية يكـــون الانحياز غير مهم، وينبغي الملاحظة هنا إلى أن العلاقة $V(\bar{Y})=\sigma_{\bar{Y}}^2$ تمثل تباينا فعليـــا. والعلاقة $V(\bar{Y})=S_{\bar{Y}}^2$ تمثل تباينا مقدرا .

2-5: حدود الثقة:

نفرض عادة أن التقديرين $\bar{Y}$ و $\hat{Y}$ متوزعان طبيعيا حول قيم المجتمع الموافقـــة. ونناقش لاحقا أسباب وآفاق مثل هذا الغرض. وإذا كان هذا الغرض قائما فإن حدود الثقة الدنيا والعليا لمتوسط المجتمع ومجموعة تكون كما يلي:

بالنسبة للمتوسط:

$$\left.\begin{aligned}\hat{\bar{Y}}_L=\bar{y}-\frac{ts}{\sqrt{n}}.\sqrt{1-f}\\ \hat{\bar{Y}}_U=\bar{y}+\frac{ts}{\sqrt{n}}.\sqrt{1-f}\end{aligned}\right\} \quad (23)$$

وبالنسبة للمجموع:

$$\left.\begin{aligned}\hat{Y}_L &= N\bar{y} - \frac{tNs}{\sqrt{n}}.\sqrt{1-f} \\ \hat{Y}_U &= N\bar{y} + \frac{tNs}{\sqrt{n}}.\sqrt{1-f}\end{aligned}\right\} \qquad (24)$$

حيث يرمز لـــ t لقيمة المتغير الطبيعي المعياري الموافقة لاحتمال الثقة المرغـــــوب والقيم الأكثر استخداما هي :

احتمال الثقة	0.50	0.80	0.90	0.95	0.99
t	0.67	1.28	1.64	1.96	2.58

وإذا كان حجم العينة أقل من 50 فيمكن الحصول على قيم t الموافقة من جــدول توزيع ستودنت بـــ (n-1) درجة من الحرية وهي درجات الحرية الموافقة لتقدير التباين S^2 أو يكون استخدام التوزيع t مشروعا تماما ، فقط عندما تتبــع الملاحظـــات y_i نفسها التوزيع الطبيعي ، ويكون حجم المجتمع N لا نهائيا وليس للميدان المعتدل عـــن التوزيع الطبيعي تأثير كبير في النتائج.

مثال: جمعت تواقيع عريضة على 676 صفحة ورق، وكل صفحة تتسع لـــ 42 توقيعا. ولكن عددا أقل من التواقيع جمع على العديد من هذه الصفحـــــات. وقـــد أحصينا عدد التواقيع في كل صفحة من صفحات عينة عشوائية حجمـــــها n = 50 (عينة مؤلفة من نحو 7% من المجتمع) وكانت النتائج كما هو مبين في الجدول التالي:

y_i أعداد التواقيع	42	41	36	32		29
f_i عدد الصفحات	23	4	1	1		1
y_i	27	23	19	16	15	14
f_i	2	1	1	2	2	1
y_i	11	10	9	7	6	5
f_i	1	1	1	1	3	2
y_i	4	3				
f_i	1	1		50		المجموع

قدر العدد الكلي للتواقيع في العريضة وضع 80% حدود ثقة لهذا التقدير

الحل: نلاحظ أن وحدة المعاينة هي الصفحة ، والملاحظة y_i هي أعداد التواقيع في كل صفحة. وبما أن نحو نصف عدد صفحات العينة يحوي العدد الأعظم من التواقيع ، أي 42 ، فقد قدمنا البيان الإحصائي على شكل جدول للتكرار ، ونلاحظ أن التوزيع يبدو كأنه في الأصل بعيد عن كونه طبيعيا. فأكبر تكرار يقع من أجل القيمة الأكبر لـ y_i ، ومع ذلك فهناك اعتقاد ، استنادا للخبرة العملية، بأن متوسطات العينات التي حجمها 50 تتوزع تقريبا وفق التوزيع الطبيعي.

حيث لدينا هنا:

$$y=\sum_i f_i\, y_i =147$$

$$n=\sum_i f_i =50$$

$$\sum_i f_i\, y_i^2 =54497$$

ومنه يكون تقدير مجموع عدد التواقيع:

$$\hat{Y}=N\bar{y}=(676)\left[\frac{1471}{50}\right]=19888$$

وبحساب تباين العينة s^2 نجد:

$$s^2=\frac{1}{n-1}\left[\sum_i f_i(y_i-\bar{y})^2\right]=\frac{1}{n-1}\left[\sum_i f_i y_i^2-\frac{\left(\sum_i f_i y_i\right)^2}{\sum f_i}\right]$$

$$=\frac{1}{49}\left[54497-\frac{(1471)^2}{50}\right]=229.0$$

ومن (24) نجد أن 80% حدود ثقة للمجموع هي:

$$\hat{Y} \pm \frac{tNS}{\sqrt{n}}.\sqrt{1-f} = 19888 \pm \frac{(1.28)(676)(15.11)}{\sqrt{50}}\sqrt{1-f}$$

حيث $f=0.074$

وهذا يعطي [18107, 21667] وهي حدود ثقة بمستوى 0.80 وعند القيام بالعدد الكامل تبين أن العدد الكلي هو في الحقيقة 21045 توقيعا.

6-2: المعاينة العشوائية مع الإعادة:

وبتطبيق أسلوب مشابه لما سبق لكن المعاينة هنا مع الإعادة ، حيث يمكن أن تظهر الوحدة i في العينة . فعندئذ:

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{N} t_i\, y_i \qquad (25)$$

وبما أن احتمال سحب الوحدة i هو $\frac{1}{N}$ وذلك عند كل عملية سحب، فالمتغير t_i يتوزع وفق التوزيع الثنائي بعدد من التكرارات يساوي n واحتمال نجاح $P=\frac{1}{N}$ ومنه

$$E(t_i) = n(\frac{1}{N}) = \frac{n}{N} \quad ; \; V(t_i) = n(\frac{1}{N})(1-\frac{1}{N}) \qquad (26)$$

وبصورة مشتركة تتبع المتغيرات t_i التوزيع الحدودي وفي هذه التوزيع:

$$CoV(t_i, t_j) = \frac{-n}{N^2} \qquad (27)$$

وباستخدام (25) و (26) نجد في حالة المعاينة مع الإعادة:

$$V(\bar{y}) = \frac{1}{n^2}\left[\sum_{i=1}^{N} y_i^2 . \frac{n(N-1)}{N^2} - 2\sum_{i<j}^{N} y_i\, y_j \frac{n}{N^2}\right]$$

$$= \frac{1}{nN}\sum_{i=1}^{N}(y_i - \bar{Y})^2 = \frac{\sigma^2}{n} = \frac{N-1}{N}.\frac{S^2}{n} \qquad (28)$$

2-6-1: إن $\bar{y}$ هو مقدر غير منحاز لـ $\bar{Y}$:

بما أن عدد العينات الممكن تشكليها في هذه الحالة هو $\binom{N+n-1}{n}$ واحتمال انتقـــاء كل منها يساوي $\frac{1}{\binom{N+n-1}{n}}$ وبالتالي يكون

$$E(\bar{y})=\sum_{K=1}^{\binom{N+n-1}{n}}\frac{\bar{y}_K}{\binom{N+n-1}{n}}=\frac{\sum_{K=1}^{\binom{N+n-1}{n}}(y_1+y_2+...+y_n)_K}{n.\binom{N+n-1}{n}}$$

ولحساب المجموع الأخير نلاحظ أنه مأخوذ على جميع العينات الممكنة في هذه الحالة، وهذا ما يؤكد لنا أن كل عنصر من عناصر المجتمع يتكرر بمقدار تكراره في هذه العينـــات. ولحســـاب عـــدد تكرارات أي عنصر نلاحظ أن عدد العينات يحوي n عنصـــرا. إذن فـــإن الجـــداء $n\binom{N+n-1}{n}$ يساوي ما تحتويه جميع هذه العينات من عناصر (مكرره وغير مكـــرره). وبما أن كل عنصر من عناصر المجتمع يتكرر بمقدار يساوي إلى تكرار أي عنصر آخـــر غيره. إذن فإن عدد تكرارات كل عنصر من عناصر المجتمع في المجمـــــوع المذكـــور يساوي $\frac{n\binom{N+n-1}{n}}{N}$ مرة ومنه نجد أن التوقع الرياضي المطلوب يصبح من الشكل :

$$E(\bar{y})=\frac{1}{n\binom{N+n-1}{n}}.\frac{n\binom{N+n-1}{n}}{N}.(y_1+y_2+.....+y_N)$$

$$=\frac{y_1+y_2+......+y_N}{N}=\bar{Y}$$

2-6-2: نتيجة:

بما أن $E(\bar{y})=\bar{Y}$ في كلتا حالتي السحب مع إعادة وبدونها، يمكننا أن نكتب

$$E(\bar{y})=\frac{\sum_{i=1}^{N}y_i}{N}\Rightarrow E=\left(\frac{\sum_{i=1}^{n}y_i}{n}\right)=\frac{\sum_{i=1}^{N}y_i}{N}$$

$$\Rightarrow E\left(\sum_{i=1}^{n}y_i\right)=\frac{n}{N}\sum_{i=1}^{N}y_i$$

مثال:

ليكن لدينا مجتمع من الدخول لست أسر في مكان ما :

38 39 43 42 41 40

والمطلوب:

1- اسحب جميع العينات الممكنة بحجم 3 عناصر في العينة الواحدة وذلك بـــدون إعادة

2- احسب متوسط الدخل في كل عينة

3- برهن أن التوقع الرياضي لمتوسطات هذه العينات يساوي متوسط المجتمع.

الحل: إن عدد العينات الممكن سحبها بدون إعادة حيث

$$N=6 \ ; \ n=3 \qquad ;\binom{N}{n}=\binom{6}{3}=\frac{6!}{3!3!}=20$$

وإن هذه العينات مع متوسطاتها معروفة في الجدول التالي:

رقم العينة	عناصر العينة			متوسط العينة
K	y_1	y_2	y_3	$\bar{y}_K$
1	40	41	42	41
2	40	41	43	41.333
3	40	41	39	40
4	40	41	38	39.666
5	40	42	43	41.666
6	40	42	39	40.333
7	40	42	38	40
8	40	43	39	40.666
9	40	43	38	40.333
10	40	39	38	39
11	41	42	43	42
12	41	42	39	40.666
13	41	42	38	40.333
14	41	43	39	41
15	41	43	38	40.666
16	41	39	38	39.333
17	42	43	39	41.333
18	42	43	38	41
19	42	39	38	40
20	42	39	38	40
				809.998

إن التوقع الرياضي لمتوسطان العينات يعطي بالعلاقة

$$E(\bar{y}_K)=\sum_{K=1}^{\binom{N}{n}}\frac{\bar{y}_K}{\binom{N}{n}}=\frac{1}{\binom{N}{n}}\sum_{K=1}^{\binom{N}{n}}\bar{y}_K$$

$$=\frac{809.998}{20}=40.4997$$

ولكن متوسط المجتمع الكلي يساوي

$$\bar{Y}=\frac{\sum_{i=1}^{N}y_i}{N}=\frac{40+41+42+43+39+38}{6}$$

$$=\frac{343}{6}=40.5$$

وبمقارنة $E(\bar{y}_K)$ مع $\bar{Y}$ نجد أن $E(\bar{y}_K)=\bar{Y}$ وذلك بإهمال الخطأ الناتج عن تقريب أرقام متوسطات العينات.

3-6-2: دراسة التماسك:

لنلاحظ جودة التقدير $\bar{y}$ بعد دراسة عدم تحيزه ، سندرس الآن تماسكه ، حيث نرى بسهوله أن هذا التقدير متماسك وذلك لأن:

$$\lim_{n\to N(\infty)} P(\bar{y}_K\to\bar{Y})=\operatorname{Lim}_{n\to N(\infty)} P\left(\frac{\sum_{i=1}^{n}y_i}{n}\to\frac{\sum_{i=1}^{N}y_i}{N}\right)=$$

وبذلك يكون $\bar{y}$ تقديرا متماسكا لـ $\bar{Y}$.

4-6-2: دراسة التباين $\sigma^2_{\bar{y}}$ أو $V(\bar{y})$:

من التعريف نجد

$$V(\bar{y})=E(\bar{y}-\bar{Y})^2=E\left[\frac{\sum_{i=1}^{n}y_i}{n}-\bar{Y}\right]^2$$

$$=E\left[\frac{\sum_{i=1}^{n}y_i-n\bar{Y}}{n}\right]^2=E\left[\frac{\sum_{i=1}^{n}(y_i-\bar{Y})}{n}\right]^2=\frac{1}{n^2}E\left[\sum_{i=1}^{n}(y_i-\bar{Y})\right]^2$$

$$=\frac{1}{n^2}E\left[\sum_{i=1}^{n}(y_i-\bar{Y})^2+\sum_{i\neq j}^{n}(y_i-\bar{Y})(y_i-\bar{Y})\right]$$

وبما أن $E(\sum_{i=1}^{n}y_i)=\frac{n}{N}\sum_{i=1}^{N}y_i$ عندئذ نجد:

$$E\left[\sum_{i=1}^{n}(y_i-\bar{Y})^2\right]=\frac{n}{N}\sum_{i=1}^{N}(y_i-\bar{Y})^2=n\sigma^2$$

أما بالنسبة للحد الأخير فنميز حالتين:

1- حالة السحب بدون إعادة: حيث نجد أن :

$$E\left[(y_i-\bar{Y})(y_j-\bar{Y})\right]=\frac{1}{N}\cdot\frac{1}{N-1}\sum_{i\neq j}^{N}(y_i-\bar{Y})(y_j-\bar{Y})$$

حيث إنه في المجموع الأخير يكون احتمال الحصول على $(y_i-\bar{Y})$ في السحبة الأولى هو الاحتمال نفسه للحصول على y_i ويساوي $\frac{1}{N}$ أما احتمال الحصول على $(y_j-\bar{Y})$, $i\neq j$ فإنه يساوي $\frac{1}{N-1}$ ، ومنه فإن احتمال الحصول على الجداء $(y_i-\bar{Y})(y_j-\bar{Y})$ يكون: $\frac{1}{N}\cdot\frac{1}{N-1}$ وهنا يجب أن نلاحظ أن المجموع أصبح في العلاقة الأخيرة مأخوذا على جميع عناصر المجتمع والتي عددها N ومن جهة أخرى نجد أن:

$$\sum_{i\neq j}^{N}(y_i-\bar{Y})(y_j-\bar{Y})=\left[\sum_{i=1}^{N}(y_i-\bar{Y})\right]^2-\sum_{i=1}^{N}(y_i-\bar{Y})^2$$

وبما أن $\sum_{i=1}^{N}(y_i-\bar{Y})=0$ عندئذ:

$$E\left[(y_i-\bar{Y})(y_j-\bar{Y})\right]=\frac{1}{N(N-1)}\left[-\sum_{i=1}^{N}(y_i-\bar{Y})^2\right]$$

ومنه نجد:

$$\sigma_{\bar{y}}^2=V(\bar{y})=\frac{1}{n^2}(n\sigma^2)-\frac{1}{n^2}\cdot\frac{1}{N-1}\sum_{i\neq j}^{n}\left[\frac{\sum_{i=1}^{N}(y_i-\bar{Y})^2}{N}\right]$$

$$=\frac{\sigma^2}{n}-\frac{1}{n^2(N-1)}\sum_{i\neq j}^{n}\sigma^2$$

وبما أن الرمز $\sum_{i\neq j}^{n}$ يعني تشكيل حدود مؤلفة من عنصرين مثل y_i , y_j فعندئ فإن عدد الحدود التي تنشأ عنه يمكن مقارنتها بعدد المتوافقات المؤلفة من عنصرين والممكن تشكيلها من n عنصراً، وبما أن الرمز السابق لا يميّز بين ترتيب العناصر (أي أنه يحوي كلا الحدين $y_i y_j$, $y_j y_i$ وهذا يعني أن كل حد يتكرر مرتين وبالتالي فإن الرمز السابق يحوي عدداً من الحدود يساوي ضعف عدد المتوافقات المؤلفة من عنصرين والتي يمكن تشكيلها من n عنصر وبذلك يكون عدد هذه الحدود مساوياً

$$2\binom{n}{2}=2\left(\frac{n!}{2!(n-2)!}\right)=n(n-1)$$

ومنه

$$\sigma_{\bar{y}}^2=\frac{\sigma^2}{n}-\frac{n(n-1)\sigma^2}{n^2(N-1)}=\frac{\sigma^2}{n}-\frac{(n-1)}{n(N-1)}\sigma^2$$

$$\Rightarrow\sigma_{\bar{y}}^2=\left(\frac{N-n}{N-1}\right)\frac{\sigma^2}{n}$$

ومن العلاقة الأخيرة نجد أنه : $\sigma_{\bar{y}}^2 \underset{n\to N}{\to} 0$

2- **وفي حالة السحب مع الإعادة** : في هذه الحالة تكون عملية سحب y_i ، y_j حوادث مستقلة ومن هذا ينتج

$$E(y_i-\bar{Y})(y_j-\bar{Y})=E(y_i-\bar{Y}).E(y_j-\bar{Y})=0$$

حيث $i \neq j$ وبالتالي فإن

$$\sigma_{\bar{y}}^2=\frac{1}{n^2}n\sigma^2=\frac{\sigma^2}{n}$$

وهنا أيضا نلاحظ أنه

$$\sigma_{\bar{y}}^2 \underset{n\to\infty}{\to} 0$$

وبالتالي نستنتج من ذلك أن فعالية التقدير $\bar{y}$ مرتبطة بحجم العينة n وتكبر كلما كبرت العينة وذلك في كلتا حالتي السحب.

2-6-5 دراسة حول تقدير إجمالي المجتمع:

نعلم أن $Y=N\bar{Y}$ يمثل إجمالي المجتمع وبما أن $E(\bar{y})=\bar{Y}$. عندئذ يمكننا ببساطة أن نقترح كتقدير لـ Y المقدر $N\bar{y}$ ونجد بسهولة

$$E(N\bar{y})=NE(\bar{y})=N\bar{Y}=Y$$

وبالتالي $\hat{Y}=N\bar{y}$ يعد تقديرا غير منحاز لإجمالي المجتمع Y. وبما أن إجمالي العينة يعطى بالعلاقة $y=n\bar{y}$ فإننا نجد أن $\hat{Y}=\frac{N}{n}y$.

حيث نجد من جهة أخرى:

$$E\left(\frac{N}{n}y\right)=\frac{N}{n}E(y)=\frac{N}{n}.\frac{n}{N}Y=Y$$

ومنه يمكننا أن نقدّر إجمالي المجتمع أيضاً بالعلاقة $\hat{Y}=\frac{N}{n}y$. وبالتالي نستنتج أن المقدارين $\frac{N}{n}y$ أو $N\bar{y}$ يعدّان مقدّرين غير منحازين لإجمالي المجتمع Y. كما أنه يمكننا بسهولة أن نبرهن على تماسك كل من هذين التقديرين وكذلك التأكد من أن فعاليتها تزداد كلما ازداد حجم العينة المسحوبة.

مثال:

ليكن لدينا مجتمع مؤلف من 100 طفل ، وسحبنا عينة عشوائية بسيطة مـــن 10 أطفال وكانت أطوالهم على الشكل التالي:

رقم الطفل	1	2	3	4	5	6	7	8	9	10
الطول بالسم C.m	60	65	70	60	70	75	65	70	75	80

والمطلوب تقدير متوسط طول الطفل في هذا المجتمع ومن ثم تقدير إجمالي الطـول في هذا المجتمع.

الحل:

نعلم أن متوسط العينة هو تقدير غير منحاز لمتوسط المجتمع عندئذ يمكن أن نقــدر متوسط هذا المجتمع بالقيمة التالية:

$$\hat{\bar{Y}}=\bar{y}=\frac{\sum_{i=1}^{n}y_i}{n}=\frac{690}{10}=69 \quad c.m$$

أما إجمالي المجتمع فيقدر بالعلاقة:

$$\hat{Y}=N\bar{y}=(100)(69)=6900 \quad c.m$$

6-6-2: دراسة حول تقدير تباين المجتمع:

كنا قد عرفنا تباين المجتمع بالعلاقة التالية:

$$\sigma^2=\frac{\sum_{i=1}^{N}(y_i-\bar{Y})^2}{N}=E(y_i-\bar{Y})^2$$

ولكن عادة في التطبيقات العملية يستخدم تعريف آخر لهذا التباين وهو:

$$S^2=\frac{\sum_{i=1}^{N}(y_i-\bar{Y})^2}{N-1}=\frac{N}{N-1}\sigma^2$$

حيث $N\neq1$.

وكذلك الأمر بالنسبة لتباين العينة حيث يعرف بـ $S^2=\frac{\sum_{i=1}^{n}(y_i-\overline{y})^2}{n-1}$ حيث $n\neq1$. وعندما $n=1 \Leftarrow S^2=0$ وهذا بدلا من تباين العينة الإحصائي المعرف بالعلاقة

$$D^2=\frac{\sum_{i=1}^{n}(y_i-\overline{y})^2}{n}$$

إن استبدال كل من D^2 بـ s^2 و σ^2 بـ S^2 ناتج عن أن تباين العينة D^2 لا يمكن أن يكون تقديرا غير منحاز لـ σ^2 وذلك لأن.

$$E(D^2)=E\left[\frac{\sum_{i=1}^{n}(y_i-\overline{y})^2}{n}\right]=\frac{1}{n}E\left[\sum_{i=1}^{n}\left[(y_i-\overline{Y})-(\overline{y}-\overline{Y})\right]^2\right]$$

$$=\frac{1}{n}E\left[\sum_{i=1}^{n}\left[(y_i-\overline{Y})^2-2(y_i-\overline{Y})(\overline{y}-\overline{Y})+(\overline{y}-\overline{Y})^2\right]\right]$$

$$=\frac{1}{n}E\left[\sum_{i=1}^{n}(y_i-\overline{Y})^2-2(y_i-\overline{Y})\sum_{i=1}^{n}(y_i-\overline{Y})+\sum_{i=1}^{n}(\overline{y}-\overline{Y})^2\right]$$

$$=\frac{1}{n}E\left[\sum_{i=1}^{n}(y_i-\overline{Y})^2-2n(\overline{y}-\overline{Y})^2+n(\overline{y}-\overline{Y})^2\right]$$

$$=\frac{1}{n}E\left[\sum_{i=1}^{n}(y_i-\overline{Y})^2-n(\overline{y}-\overline{Y})^2\right]$$

$$=\frac{1}{n}\sum_{i=1}^{n}E(y_i-\overline{Y})^2-nE(\overline{y}-\overline{Y})^2$$

$$=\frac{1}{n}\left[\sum_{i=1}^{n}\sigma^2-n\sigma_{\overline{y}}^2\right]=\frac{1}{n}\left[n\sigma^2-n\sigma_{\overline{y}}^2\right]$$

$$E(D^2)=\sigma^2-\sigma_{\overline{y}}^2$$

وبالتالي فإن $E(D^2)$ لا يساوي σ^2 وهذا يعني أن D^2 لا يمثل تقديـــرا غـــير منحاز لـــ σ^2. ومنه نميز الآن بين حالتين:

1- حالة السحب بدون إعادة: رأينا في فقرة سابقة أن:

$$\sigma_{\bar{y}}^2=\frac{N-n}{N-1}.\frac{\sigma^2}{n}$$ وبالتعويض $E(D^2)$

نجد:

$$E(D^2)=\sigma^2-\frac{N-n}{n(N-1)}.\sigma^2=\left[\frac{n(N-1)-N+n}{n(N-1)}\right]\sigma^2$$

$$\Rightarrow E(D^2)=\frac{N(n-1)}{n(N-1)}.(\sigma^2)$$

2- حالة السحب مع الإعادة : رأينا أن $\sigma_{\bar{y}}^2=\frac{\sigma^2}{n}$ ومنه

$$E(D^2)=\sigma^2-\frac{\sigma^2}{n}=\frac{n-1}{n}\sigma^2$$

وبالتالي إذا أخذنا $s^2=\frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{n-1}$ فإن

$$s^2=\frac{nD^2}{n-1}\Rightarrow\ E(s^2)=E\left(\frac{nD^2}{n-1}\right)=\frac{n}{n-1}.\frac{n-1}{n}\sigma^2$$

$$\Rightarrow\ E(S^2)=\sigma^2$$

حيث نجد بذلك أن s^2 هو مقدّر غير منحاز لـــ σ^2. وهذا مسوغ، في حالـــة السحب مع الإعادة باعتماد s^2 بدلاً من D^2.

وفي حالة السحب بدون إعادة ، نجد أنه حتى يكون D^2 تقديراً غير منحاز لـــــ σ^2 يجب ضربه بمعامل تصحيح من الشكل: $\left(\frac{n}{n-1}\right)\left(\frac{N-1}{N}\right)$ مما سيضطرنا لوضـــع تعريف ثالث لتباين العينة يكون مرتبطاً بالعدد N والذي ليس له علاقة بالعينة أساساً.

مثال:

ليكن لدينا مجتمع مؤلف من خمسة قرى e, d, c, b, a, ويراد تقدير متوسط عـــدد سكان القرية في ذلك المجتمع بسحب عينة عشوائية بسيطة مؤلفة من ثلاث قـــرى. لذلك قام الباحثون بسحب ثلاث قرى بدون إعادة ولتكن a, d, c, وأحصـــوا عـــدد سكان كل منها فكان كما يلي : 550 ,625 ,700 نسمة والمطلوب الآن:

1- إيجاد تقدير متوسط عدد سكان القرية في ذلك المجتمع وإجمالي السكان.

2- إيجاد تقدير لتباين المجتمع.

3- لنفترض الآن أن عدد سكان القرى الخمس معلوم ويساوي على الـــترتيب:
600, 625, 550, 500, 700

والمطلوب:

a- حساب عدد العينات الممكن سحبها بدون إعادة والمؤلفة من ثلاث قرى.

b- سحب هذه العينات الممكنة وحساب متوسط كل عينة.

c- حساب تباين كل من هذه العينات.

d- التأكد من أن التوقع الرياضي لمتوسطات هذه العينات يساوي متوسط المجتمع.

e- التأكد من أن التوقع الرياضي لتباينات تلك العينات يساوي تباين المجتمع.

4- أعد كل الحسابات المطلوبة في الطلب (3) في حالة السحب مع الإعادة.

الحل:

1- من المعلوم أن متوسط عدد سكان القرية في ذلك المجتمع يمكن تقديره بوساطة متوسط العينة المسحوبة، لذلك نجد أن

$$\hat{\bar{Y}} = \bar{y} = \frac{700+625+500}{3} = 62$$

وبالتالي فإن تقدير إجمالي عدد السكان يساوي

$$\hat{Y} = N\bar{y} = 5 \times 625 = 3125$$

2- إن تقدير تباين المجتمع يحسب بوساطة تباين العينة s^2

$$s^2=\frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{n-1}=\frac{1}{2}\left[(-75)^2+(0)^2+(75)^2\right]$$
$$=5625$$

إن قيمة s^2 هذه تعد تقديرا غير منحاز لتباين المجتمع s^2 وذلك لأن السحب جرى بدون إعادة.

3- (a) نعلم أن عدد العينات التي يمكن سحبها في حالة السحب بدون إعادة يساوي $\binom{N}{n}$ أي:

$$\binom{5}{3}=\frac{5!}{3!2!}=1$$

b ,c- يمكننا مباشرة الإجابة عن الطلبين من خلال الجدول التالي:

رقم العينة K	عدد سكان عناصر العينة y_i	متوسط العينة $\bar{y}_i$	$s_i^2=\frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{n-1};n=3$: (تباين العينة)
1	700 500 550	583	10833.5
2	700 500 625	608	10208.5
3	700 500 600	600	10000
4	700 550 625	625	5625
5	700 550 600	617	5833.5
6	700 625 600	642	2708.5
7	500 550 625	558	3958.5
8	500 550 600	550	2500
9	500 625 600	575	4375

10	550 625 600	592	1458.5
المجموع		5950	57501

d- وللتأكد من أن التوقع الرياضي لمتوسطات هذه العينات يســـاوي متوســط المجتمع، نحسب أولا متوسط المجتمع

$$\overline{Y}=\frac{\sum_{i=1}^{N}y_i}{N}=\frac{700+600+500+550+625}{5}=595$$

وبحساب التوقع الرياضي لــ $\overline{y}$ نجد أن:

$$E(\overline{y})=\frac{1}{\binom{N}{n}}\sum_{K=1}^{\binom{N}{n}}\overline{y}_K=\frac{5950}{10}=595$$

ومنه نجد أن $E(\overline{y})=\overline{Y}$

e- وللتأكد من أن التوقع لتباينات العينات يساوي تباين المجتمع، نحسب أولا تباين المجتمع S^2 وذلك لأن السحب جرى بدون إعادة فنجد أن:

$$S^2=\frac{1}{N-1}\sum_{i=1}^{N}(y_i-\overline{Y})^2=\frac{1}{4}[(105)^2+(-95)^2+(-45)^2$$
$$+(30)^2+(5)^2]=5750$$

ولكن توقع تباينات العينات يساوي

$$E(s^2)=\frac{1}{\binom{N}{n}}\sum_{i=1}^{\binom{N}{n}}s_i^2=\frac{57501}{10}=57501$$

وبالمقارنة نجد تقريبا أن $E(s^2)=S^2$

4- إن معالجة الطلب الرابع لا تختلف كثيرا عما سبق وبخاصة إذا علمنا أن عـــدد العينات الممكنة يكون:

$$\binom{N+n-1}{n}=\frac{(N+n-1)!}{n!(N-1)!}=\frac{7!}{3!4!}=35$$

وإن تباين المجتمع يجب أن يحسب من العلاقة

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{Y})^2$$

وبإجراء الحسابات اللازمة فإننا سنحصل على النتائج السابقة نفسها. ونترك أمر هذه الحسابات للطالب على سبيل التمرين.

7-2: تقدير تباين متوسط العينة المسحوبة $\sigma_{\bar{y}}^2$ وتباين إجمالي المجتمع $\sigma_{\hat{y}}^2$:

a- لإيجاد تقدير كل من هذين التباين نميز حالتين:

1- حالة السحب بدون إعادة: حيث رأينا في فقرة سابقة

$$\sigma_{\bar{y}}^2 = \frac{N-n}{N-1}.\frac{\sigma^2}{n}$$

وكذلك

$$\sigma^2 = \frac{N-1}{N}S^2$$

وبالتالي:

$$\sigma_{\bar{y}}^2 = \frac{N-n}{N-1}.\frac{N-1}{N}.\frac{S^2}{n} = \frac{N-n}{N}.\frac{S^2}{n}$$

وبما أن S^2 هو تقدير غير منحاز لـ S^2 فإن تقدير $\sigma_{\bar{y}}^2$ سيعطى بالعلاقة:

$$\hat{\sigma}_{\bar{y}}^2 = \frac{N-n}{N}.\frac{S^2}{n}$$

2- حالة السحب مع الإعادة: حيث نجد من العلاقة $\sigma_{\bar{y}}^2 = \frac{\sigma^2}{n}$ بما أن s^2 هو تقدير غير منحاز لـ σ^2، عندئذ يمكننا بساطة أن نقدر $\sigma_{\bar{y}}^2$ في هذه الحالة بالعلاقة: $\hat{\sigma}_{\bar{y}}^2 = \frac{s^2}{n}$ ويمكننا أيضا التأكد من أن كلا من التقديرين السابقين هو تقدير غير منحاز حيث نجد أن:

في حالة السحب بدون إعادة:

$$E(\hat{\sigma}_{\bar{y}}^2) = E\left[\frac{N-n}{N}.\frac{s^2}{n}\right] = \frac{N-n}{nN}Es^2 = \frac{N-n}{nN}S^2$$

وفي حالة السحب مع الإعادة:

$$E(\hat{\sigma}_{\bar{y}}^2)=E(\frac{s^2}{n})=\frac{1}{n}E(s^2)=\frac{\sigma^2}{n}=\sigma_{\bar{y}}^2$$

b- أما فيما يتعلق بتقدير إجمالي المجتمع Y فنعالجه كما يلي:

نحن نعلم أن إجمالي المجتمع الحقيقي يعطي بالعلاقة:

$$Y=N\bar{Y}\Rightarrow\hat{Y}=N\bar{y}$$

$$\sigma_{\hat{Y}}^2=E[\hat{Y}-Y]^2=E[N\bar{y}-N\bar{Y}]^2$$

$$=N^2E(\bar{y}-\bar{Y})^2=N^2\sigma_{\bar{y}}^2$$

وهنا نميز حالتين:

1- حالة السحب بدون إعادة: نعوض $\sigma_{\bar{y}}^2$ بما يساويه

$$\sigma_{\hat{y}}^2=N^2.\frac{N-n}{N-1}.\frac{\sigma^2}{n}=\frac{N^2.(N-n)(N-1)}{(N-1).N}.\frac{S^2}{n}$$

وبما أن s^2 هو تقدير غير منحاز لــ S^2. عندئذ يكون تقدير تباين إجمالي المجتمع معطى بالعلاقة التالية:

$$\hat{\sigma}_{\hat{y}}^2=N^2.\frac{N-1}{N}.\frac{s^2}{n}$$

2- وفي حالة السحب مع الإعادة: هنا لدينا

$$\hat{\sigma}_{\hat{y}}^2=N^2.\frac{s^2}{n}$$

مثال:

بالعودة للمثال السابق: لنبحث عن تقدير لتباين متوسط العينة المسحوبة $\sigma_{\bar{y}}^2$ ولتباين إجمالي المجتمع $\sigma_{\hat{y}}^2$

الحل:

لدينا

$$\hat{\sigma}_{\bar{y}}^2=\frac{N-n}{N}.\frac{s^2}{n}=\frac{5-3}{5}.\frac{5625}{3}=750$$

وكذلك يمكننا أن نحسب $\hat{\sigma}_{\bar{y}}^2$ لجميع العينات العشر الواردة في الجدول ونتـأكد من أن التوقع الرياضي للتقديرات $\hat{\sigma}_{\bar{y}}^2$ يساوي $\sigma_{\bar{y}}^2$ نفسها حيث:

$$\sigma_{\bar{y}}^2 = \frac{5-3}{5} . \frac{5750}{3} = 767 \;\; ; \; (S^2 = 5750 \text{ حيث})$$

أما بالنسبة لتباين إجمالي المجتمع فيمكننا أن نحسبه من:

$$\hat{\sigma}_{\bar{y}}^2 = N^2 \, \hat{\sigma}_{\bar{y}}^2 = (25)(750) = 1875$$

هذا ويمكننا أيضا أن نحسب التقديرات $\hat{\sigma}_{\bar{y}}^2$ لجميع العينات العشـر الـواردة في الجدول ونتأكد من أن التوقع الرياضي لهذه التقديرات يساوي $\sigma_{\hat{y}}^2$ حيث

$$\hat{\sigma}_{\bar{y}}^2 = N^2 \, \sigma_{\bar{y}}^2 = (25)(767) = 19175$$

وهذا يعني أنه يمكننا أن نحصل على تقدير $\sigma_{\hat{y}}^2$ وذلك بضرب $\hat{\sigma}_{\bar{y}}^2$ بـ N^2 وأن هذه العلاقة تفيدنا جدا في الحسابات العملية.

8-2: تقدير الخطأ المعياري للتقديرات:

1- حالة السحب بدون إعادة:

إن الخطأ المعياري لمتوسط العينة يعطى بالعلاقة التالية:

$$\hat{\sigma}_{\bar{y}} = \sqrt{\frac{N-n}{nN}} . s.$$

وإن الخطأ المعياري لإجمالي المجتمع يعطى بالعلاقة التالية:

$$\hat{\sigma}_{\hat{Y}} = N . \sqrt{\frac{N-n}{nN}} . s = N \hat{\sigma}_{\bar{y}}$$

2- حالة السحب مع الإعادة :

إن الخطأ المعياري لمتوسط العينة

$$\hat{\sigma}_{\bar{y}} = \frac{s}{\sqrt{n}}$$

وكذلك

$$\hat{\sigma}_{\hat{Y}} = N . \frac{s}{\sqrt{n}}$$

9-2: تقدير مدى الثقة في التقدير:

نحن دائما نرغب أن تكون صحة التقديرات التي نحصل عليها لوسطاء المجتمع ذات احتمال كبير وبالتحديد أكبر من 0.95. ففي حالة $\bar{y}$. مثلا والتي تعد تقديرا لـ $\bar{Y}$ يجب أن يتحقق لدينا العلاقة التالية:

$$P[|\bar{y}-\bar{Y}|<{}^{\varepsilon}\beta]=\beta>0.95$$

أي يجب أن نحصل على مجال نصف طوله ε_β يعطي المقدار $\bar{Y}$ باحتمال قدره β ، حيث β أكبر من 0.95 وأن ε_β عدد موجب بتبديل ε_β و $Z_\beta\,\hat{\sigma}_{\bar{y}}$ نجد

$$P[|\bar{y}-Y|<Z_\beta\,\hat{\sigma}_{\bar{y}}]=\beta\Rightarrow$$

$$P[-Z_\beta\,\hat{\sigma}_{\bar{y}}<\bar{y}-\bar{Y}<Z_\beta\,\hat{\sigma}_{\bar{y}}]=\beta\Rightarrow$$

$$P[\bar{y}-Z_\beta\,\hat{\sigma}_{\bar{y}}<\bar{Y}<\bar{y}+Z_\beta\,\hat{\sigma}_{\bar{y}}]=\beta$$

وهذا يعني أن احتمال أن يغطي المجال

$$P[\bar{y}-Z_\beta\,\hat{\sigma}_{\bar{y}}\,,\bar{y}+Z_\beta\,\hat{\sigma}_{\bar{y}}]$$

المقدار $\bar{y}$ وهو β .

وإذا كانت $\bar{y}$ خاضعة للتوزيع الطبيعي (كما نفترض) فلقد رأينا من خلال جدول التوزيع الطبيعي المعياري أنه:

إذا كان $Z_\beta=1$ فإن P= 0.68 والعكس بالعكس

وإذا كان $Z_\beta=2$ فإن P= 0.96 والعكس بالعكس

وإذا كان $Z_\beta=3$ فإن P = 0.997 والعكس بالعكس

فإن أردنا الحصول على مجال يغطي $\bar{y}$ باحتمال قدره 0.99 فإنه علينا أن نفترض أن Z = 3 وعندها نحصل على المجال.

$$P[\bar{y}-3\hat{\sigma}_{\bar{y}}\ \ ,\bar{y}+3\hat{\sigma}_{\bar{y}}]$$

حيث ندعوه بمدى الثقة أو مجال الثقة. أما قيمة الاحتمال الموافقة فتسمى احتمال الثقة أو معامل احتمال الثقة.

وكذلك يمكننا أن نعرّف مدى الثقة لتقدير إجمالي المجتمع حيث أننا نستبدل في العلاقات السابقة $\hat{\sigma}_{\bar{y}}$ بمقابلها $\hat{\sigma}_{\hat{Y}}$ ومنه نجد أن المجال سيكون من الشكل التالي:

$$[\hat{Y}-Z_{\beta}\,\hat{\sigma}_{\hat{Y}}\,,\hat{Y}+Z_{\beta}\,\hat{\sigma}_{\hat{Y}}]$$

والذي يغطي Y باحتمال قدره β .

مثال:

ليكن لدينا مجتمع مؤلف من N = 2000 طالب ، ولنفرض أننا سحبنا عينه بدون إعادة ذات حجم n = 100 وكان متوسط استهلاك الحليب اليومي في هذه العينة $\bar{y}=0.8$ كثيراً وكان تباين الاستهلاك في العينة $s^2=0.095$. والمطلوب: تحديد مدى الثقة لمتوسط المجتمع ولإجمالي المجتمع لكمية الحليب المستهلكة وذلك باحتمال قدره 0.95 و0.99 لكل منها على الترتيب.

الحل:

1- بالنسبة لمتوسط المجتمع نعلم أن مدى الثقة يعطى بالعلاقة:

$$\bar{y}-Z\,\hat{\sigma}_{\bar{y}} \quad <\bar{Y}<\bar{y}+Z\,\hat{\sigma}_{\bar{y}}$$

ونجد Z = 2 يقابل الاحتمال P = 0.95 فإن:

$$0.8-(2)(0.03) <\bar{Y}<0.8+(2)(0.03)$$

حيث أن $\hat{\sigma}_{\bar{y}}=\sqrt{\frac{N-n}{N}.\frac{s^2}{n}}=0.03$

وبالتالي وبثقة 0.95 يكون متوسط الاستهلاك اليومي للحليب بالنسبة للطالب الواحد محصوراً في المجال [0.74 , 0.86].

وإن الأمر لا يختلف كثيراً من أجل تحديد مدى الثقة الإجمالي للاستهلاك حيث إن

$$\hat{Y}-Z\,\hat{\sigma}_{\hat{Y}} \;<Y<\hat{Y}+Z\,\hat{\sigma}_{\hat{Y}}]$$

$$\hat{Y}=N\bar{y}=(2000)(0.8)=1600$$

$$\hat{\sigma}^2_{\hat{Y}}=N^2.\frac{N-n}{N}.\frac{S^2}{n}=(2000)^2.\left(\frac{2000-100}{2000}\right).\left(\frac{0.095}{100}\right)$$

$$=3420 \Rightarrow \;\hat{\sigma}_{\hat{Y}}=\sqrt{3420}=58.5$$

وبذلك يكون مدى الثقة الإجمالي الاستهلاك بثقة 99% (أي Z=3):

$$1600-3(58.5) < Y < 1600+3(58.5)$$

ومن نكون واثقين 99% من أن إجمالي استهلاك الحليب سيكون محصورا في المجال [1224.5 , 1775.5].

10-2: تقدير ثابت التباين النسبي لـــ $\bar{y}$:

يعرف ثابت التباين النسبي بالعلاقة: $C^2 = \dfrac{\sigma_{\bar{y}}^2}{\bar{y}^2}$

فإن بدلنا بـــ $\sigma_{\bar{y}}^2$ قيمتها من العلاقة السابقة سنحصل عندئذ على تقديـــر غـــير منحاز لـــ C:

1- في حالة السحب بدون إعادة:

$$\hat{C}^2 = \frac{N-n}{nN} \cdot \frac{s^2}{\bar{y}^2}$$

2- وفي حالة السحب مع الإعادة:

$$\hat{C}^2 = \frac{s^2}{n\bar{y}^2}$$

وإن ثابت التباين يفيدنا في تقدير فعالية التقديرات التي حصلنا عليها وذلـــك إذا علمنا أن قيمة C المقبولة يجب أن لا تزيد على 0.10 ولا فإن التقديرات التي حصلنـــا عليها تكون غير مقبولة.

11-2: تقدير الدقة وحجم العينة:

إن حجم العينة المطلوبة يتعلق بالمؤشر المراد حسابه لذلك فإننا سندرس لإمكـــان تقديره حسب المؤشرات المختلفة.

1- بالنسبة لـــ $\bar{y}$:

كنا قد ذكرنا سابقا أن الدقة يمكن أن تعرف بـــ $d = Z\sigma_{\bar{y}}$ وبتعويـــض $\sigma_{\bar{y}}$ بقيمتها التي وجدنا في العلاقات السابقة، ففي حالة السحب بدون إعادة نجد :

$$\hat{d}^2 = Z^2 \frac{N-n}{nN} s^2$$

وفي حالة السحب مع الإعادة نجد:

$\hat{d}^2 = Z^2 \frac{s^2}{n}$. أما تقدير حجم العينة فنحصل عليه من هاتين المعادلتين كالتـــالي:

في حالة السحب بدون إعادة نجد:

$$\hat{n} = \frac{NZ^2 s^2}{N\hat{d}^2 + Z^2 s^2}$$

وفي حالة السحب مع الإعادة نجد:

$$\hat{n}_0 = \frac{Z^2 s^2}{\hat{d}^2}$$

2- بالنسبة لإجمالي المجتمع:

حيث إن الدقة معرّفة بالعلاقة $\hat{d}^2 = Z^2\, \hat{\sigma}^2_{\hat{Y}}$. ففي حالة السحب بدون إعادة:

$$\hat{d}^2 = \left(\frac{N^2(N-n)}{nN} s^2\right) Z^2 = \frac{N(N-n)}{n} Z^2 s^2$$

ويكون الحجم المطلوب:

$$\hat{n} = \frac{N^2 Z^2 s^2}{\hat{d}^2 + NZ^2 s^2}$$

وفي حالة السحب مع الإعادة نجد:

$$\hat{d}^2 = Z^2 . N^2 . \frac{s^2}{n}$$

ويكون الحجم المطلوب: $\hat{n}_0 \frac{N^2 Z^2 s^2}{\hat{d}^2}$

ويمكننا تقدير الدقة النسبية بناء على تقدير ثابت التباين النسبي حيث:

$$d'^2 = \frac{Z^2 \sigma^2_{\bar{y}}}{\bar{y}^2} \quad \text{أو} \quad d'^2 = \frac{d^2}{\bar{y}^2}$$

أي $d'^2 = Z^2 C^2$ وبذلك يكون تقدير الدقة النسبية كتالي: $\hat{d}'^2 = Z^2 \hat{C}^2$ وبالتالي:

في حالة السحب بدون إعادة نجد:

$$\hat{d}'^2 = Z^2 . \frac{N-n}{nN} . \frac{s^2}{\bar{y}^2} = \frac{Z^2 (N-n)}{nN} C_1^2$$

حيث فرضنا أن $C_1^2=\frac{S^2}{\bar{y}^2}$ وبالتالي حجم العينة يكون:

$$\hat{n}=\frac{NC_1^2 Z^2}{N\hat{d}'^2+Z^2C_1^2}$$

وفي حالة السحب مع الإعادة نجد:

$$\hat{d}'^2=Z^2.\frac{s^2}{n\bar{y}^2}=\frac{Z^2C_1^2}{n}$$

وبالتالي فإن تقدير حجم العينة يكون:

$$\hat{n}_0=\frac{N^2C_1^2}{\hat{d}'^2}$$

ومن $\hat{n}$ و $\hat{n}_0$ لدينا:

$$N\hat{n}_0=\frac{NZ^2C_1^2}{\hat{d}'^2};N+\hat{n}_0=\frac{Z^2C_1^2}{\hat{d}'^2}+N$$

$$\Rightarrow N+\hat{n}_0=\frac{Z^2C_1^2+N\hat{d}'^2}{\hat{d}'^2}$$

ومن العلاقتين الأخيرتين نجد:

$$\frac{N\hat{n}_0}{N+\hat{n}_0}=\frac{\frac{NZ^2C_1^2}{\hat{d}'^2}}{\frac{Z^2C_1^2+N\hat{d}'^2}{\hat{d}'^2}}=\hat{n}$$

ومنه نجد:

$$\hat{n}=\frac{N\hat{n}_0}{N+\hat{n}_0} \quad \Rightarrow \quad =\hat{n}<\hat{n}_0$$

ومنه تكون $\hat{n}_0$ أكبر من $\hat{n}$ دائما. وعلاقتها تعطى بالعلاقة $\hat{n}=\frac{N\hat{n}_0}{N+\hat{n}_0}$ لجميع التقديرات.

مثال:

عين حجم العينة الواجب سحبها بدون إعادة من مجتمع الأبقار لتقدير وسطي وزن البقرة مع العلم أن N= 2000 والتباين المقدر $S^2=600$ وبدقة d= 5 وفي حالة Z = 3.

الحل:

لدينا

$$\hat{n}=\frac{NZ^2s^2}{N\hat{d}^2+Z^2s^2}=\frac{(2000)(3)^2(600)}{(2000)(5)^2+(3)^2(600)}=195$$

ولو فرضنا أن N = 20000 فإن حجم العينة المطلوب

$$\hat{n}=214$$

وإذا كان المجتمع كبيرا (غير محدود) فإننا نستخدم علامة حجم العينة في حالة السحب مع الإعادة حيث نجد أن:

$$\hat{n}=\frac{Z^2S^2}{\hat{d}^2}=\frac{(9)(600)}{25}=216$$

وبالتالي هنا نلاحظ أن تغيير في حجم المجتمع N لا يؤثر بشكل ملحوظ في حجم العينة فيما إذا بلغ هذا الحجم حدا كبيرا حيث نرى أن:

$N = 2000 \Rightarrow n = 195$
$N = 20000 \Rightarrow n = 214$
$N = \infty \Rightarrow n = 216$

ونستنتج هنا أن نسبة حجم العينة في المجتمع الكبير:

$$\frac{n}{N}=\frac{214}{20000}=0.01$$

تكون كافية لتحديد معالم هذا المجتمع ضمن الشروط السابقة.

12-2: تقدير نسبة خاصة معينة في المجتمع :

كثيرا ما نحتاج في البحوث العلمية وبخاصة الاقتصادية أو الاجتماعية إلى تقدير نسبة خاصة ما في المجتمع. فمثلا نسبة عدد المدخنين في مدينة ما أو بنسبة عدد الطلاب فيها إلى المجتمع الكلي أو نسبة الأفراد العاطلين عن العمل أو ذوي الدخل

المحدود أو نسبة الأفراد ذوي الدخل الفاحش أو نسبة مالكي الأراضي والعقـارات في منطقة ما. فلدراسة مثل هذه الحالات يقسم المجتمع إلى قسمين متنافيين تماما. القسـم الأول يتألف من العناصر التي تتصف بتلك الخاصة المفروضة والقسم الثاني يتألف مـن العناصر التي لا تتصف بتلك الخاصة. فإذا رمزنا لعدد عناصر المجتمع بـ N ولعـدد عناصر القسم الأول A فإننا نجد أن نسبة عناصر المجتمع الذين يتصفـون بالخاصـة المفروضة هي: $R=\frac{A}{N}$

وأن نسبة عناصر المجتمع التي لا تتصف بالخاصة

$$Q=\frac{N-A}{N}=1-R$$

فلتقدير R أو Q في المجتمع نستخدم النسب المقابلة لها في العينة المسحوبة، فـإذا رمزنا لحجم العينة بـ n ولعدد العناصر من العينة التي تتصف بالخاصة المفروضة بـ a فإن نسبة العناصر المتصفة بالخاصة من العينة تكون $r=\frac{a}{n}$ ونسبة العنـاصر غـير المتصفة بالخاصة من العينة تكون: $q=\frac{n-a}{n}=1-r$.

لنبحث الآن فيما إذا كانت النسبة r والتي تم الحصول عليها بوساطة العينة تمثـل التقدير غير المنحاز والمتماسك والفعال للنسبة الأصلية R في المجتمع. من أجل ذلـك تعد $y_1,y_2,.......y_N$ تمثل عناصر المجتمع ولنأخذ $y_i=1$ إذا كان متصفا بالخاصـة المفروضة و $y_i=0$ إذا كان غير متصف بالخاصة المفروضـة وذلـك مـن أجـل $i=1,2,.......N$. ومنه يكون $A=\sum_{i=1}^{N}y_i$

وكذلك

$$\frac{\sum_{i=1}^{N}y_i}{N}=\frac{A}{N}=R$$

وهذه ما هي إلا عبارة المتوسط الحسابي للمجتمع وهي نفسها نسـبة عنـاصر المجتمع المتصفة بتلك الخاصة.

وكذلك بالنسبة للعينة حيث نفترض أيضا أن عناصر العينة y_i تـــأخذ إحـــدى القيمتين 1 أو 0 وذلك حسب اتصافها بتلك الخاصة أو عدم اتصافها بها على الترتيب. ولكن هنا نجد أن:

$$r=\frac{\sum_{i=1}^{n}y_i}{n}=\bar{y}$$

وهي تمثل عبارة المتوسط الحسابي للعينة. وبرهنا سابقا على أن المتوسط الحســابي للعينة يمثل تقديرا غير منحاز للمتوسط الحسابي للمجتمع وبالتالي r هي تقدير غـــير منحاز لـــ R.

أما بالنسبة لدراسة مؤشرات النسبة الأخرى فإن:

1- تباين المجتمع: يعرف بـــ

$$S^2=\frac{1}{N-1}\sum_{i=1}^{N}(y_i-\bar{Y})^2=\frac{1}{N-1}\left[\sum_{i=1}^{N}y_i^2-N\bar{Y}^2\right]$$

وبما أن y_i تأخذ القيم 1 أو 0 فإن

$$\sum_{i=1}^{N}y_i^2=A=NR \quad \text{و} \quad \bar{Y}^2=R^2$$

(حيث أن : $\bar{Y}=\frac{\sum_{i=1}^{N}Y_i}{N}=\frac{NR}{N}=R$) وبالتالي نجد:

$$s^2=\frac{1}{N-1}(NR-NR^2)=\frac{N}{N-1}R(1-R)$$

أي أن :

$$S^2=\frac{N}{N-1}RQ$$

2- تباين العينة: يعطى بشكل مماثل بالعلاقة

$$s^2=\frac{n}{n-1}rq$$

ويمكننا أن نبرهن على أن s^2 يعد تقديرا غير منحاز لـــ S^2 وذلـــك باتبـــاع المراحل السابقة نفسها في البرهان.

3- تقدير تباين r:

في حالة السحب بدون إعادة نجد:

$$\hat{\sigma}_r^2 = \frac{N-n}{N} \cdot \frac{s^2}{n}$$

وفي حالة السحب مع الإعادة نجد:

$$\hat{\sigma}_r^2 = \frac{s^2}{n}$$

ومنه نجد أن تقدير تباين r في حالة السحب بدون إعادة:

$$\hat{\sigma}_r^2 = \frac{N-n}{N} \cdot \frac{1}{n} \cdot \frac{n}{n-1} r.q = \frac{N-n}{N(n-1)} r.q$$

وفي حالة السحب مع الإعادة نجد:

$$\hat{\sigma}_r^2 = \frac{s^2}{n} = \frac{n}{n(n-1)} r.q = \frac{rq}{n-1}$$

4- الخطأ المعياري ومدى الثقة:

يعرفـه الخطأ المعياري للنسبة r بالمقدار $\hat{\sigma}_r$ ويكون بالتالي مدى الثقة:

$$r - Z\hat{\sigma}_r \quad < R < r + Z\hat{\sigma}_r$$

وبفرض أن r لها توزيع طبيعي وهذا طبعا مرتبط بحجم العينة ، فيمكننا أن نحصل على قيم Z الموافقة للاحتمال المطلوب، كما رأينا سابقا عندما أجرينا هذه المناقشـــة على $\bar{y}$ ومن ثم نحدد مدى الثقة لتقدير R.

5- الدقة وحجم العينة:

تعرف الدقة في هذه الحالة بـــ $d^2 = Z^2 \hat{\sigma}_r^2$ ومنها نجد أنه في حالة السحب بدون إعادة

$$d^2 = Z^2 \cdot \frac{N-n}{N} \cdot \frac{rq}{n-1} \approx Z^2 \cdot \frac{N-n}{nN} r\dot{q}$$

وفي حالة السحب مع الإعادة.

$$d^2=\frac{Z^2rq}{n-1}\approx\frac{Z^2rq}{n}$$

ومن العلاقتين السابقتين يمكننا أن نجد حجم العينة الذي يحقق لنا الدقة المطلوبة حيث يكون لدينا في حالة السحب بدون إعادة:

$$n=\frac{NZ^2rq}{Nd^2+Z^2rq}$$

وفي حالة السحب مع الإعادة:

$$n=\frac{Z^2\,rq}{d^2}$$

ونرفق الآن جدولا إحصائيا يبين حجم العينة المطلوب لكي يكون توزيع r توزيعا طبيعيا وذلك حسبا تكون قيمة r في المجتمع المدروس.

حجم العينة المطلوب n	النسبة r
$n\geq 30$	$r=0.5$
$n\geq 50$	$r=0.6$ أو $r=0.4$
$n\geq 80$	$r=0.7$ أو $r=0.3$
$n\geq 200$	$r=0.8$ أو $r=0.2$
$n\geq 600$	$r=0.9$ أو $r=0.1$
$n\geq 1400$	$r=0.95$ أو $r=0.05$

مثال:

عند تقدير نسبة العائلات التي تمتلك مذياعين أو أكثر في مدينة ما، كان لدينا N= 8000 عائلة وسحبنا عينة عشوائية n = 101 فوجدنا أن r = 0.6 احسب مدى الثقة لـ R من أجل ثقة 0.99.

الحل :

إن $r-3\hat{\sigma}_r \leq R\leq r+3\hat{\sigma}_r$

يمثل مدى ثقة بمستوى 99%. ولكن:

$$\hat{\sigma}_r^2 = \frac{N-n}{N} \cdot \frac{rq}{n-1} = \frac{8000-101}{8000} \cdot \frac{(0.6)(0.4)}{101-1} = 0.002368$$

$$\Rightarrow \hat{\sigma}_r = 0.0486$$

ومنه

$$0.6-3(0.0486) \leq R \leq 0.6+3(0.0486)$$

$$0.464 \quad \leq R \leq 0.7358$$

أي نكون واثقين بمقدار 0.99 من أن النسبة الحقيقية R للعائلات التي تمتلك مذياعين أو أكثر في المدينة لن تقل عن 0.464 ولن تزيد على 0.7358.

ومن أجل $n = 400$ نجد أن $0.53 < R < 0.67$

ومن أجل $n = 1000$ نجد أن $0.56 < R < 0.65$

وهكذا نرى مقدار تأثير حجم العينة في تضييق مدى الثقة.

13-2: دراسة تقدير المعدلات:

إننا في كثير من الأحيان وفي البحوث الإحصائية نضطر إلى تعريف وحدة المعاينة بشكل يضم داخلها جملة من العناصر الأولية (واحدة لا يراد تجزئتها كالأسرة مثلا) ونضطر بالمقابل أيضا أن نجيب عن عدد من الأسئلة المتعلقة بتلك العناصر الأولية الداخلية. فمثلا عند إجراء معاينة على أسر مدينة ما فإننا يجب أن نحصل على المعلومات المطلوبة عن تلك الأسر ولكن غالبا ما يطلب منا أن نحسب معدل نصيب الفرد من دخل الأسرة في المجتمع ككل وذلك بحسابه من معلومات العينة. أو يطلب منا أيضا معدل نصيب الفرد من الإنفاق على الأغذية ، أو قد يهمنا معرفة معدل نصيب الفرد من المساحة السكنية في المجتمع المدروس ، مهننا لابد من حساب مجموع الدخول أو الاتفاقات أو المساحات لجميع الأسر في العينة ثم تقسيمها على إجمالي عدد أفراد أسر تلك العينة وذلك حتى نحصل على تقدير لمعدل نصيب الفرد في ذلك المجتمع من الدخل أو الإنفاق أو المساحة. ومسألة تقدير المعدلات يمكن مواجهتها في مختلف مجالات الحياة على أن يكون المجتمع المدروس متجانسا بالنسبة لقيم الخاصة التي نريد حساب معدلها وذلك حتى لا نقع بخطأ في التقدير ونشوه الوقائع.

فإذا كنا نريد تقدير نصيب الفرد من المساحة السكنية مثلا، فإنه يجب علينا أن نختار مجتمعا متجانسا من حيث تلك المساحة، وإلا فعلينا أن نبوب واحدات المعاينة

في مجموعات متجانسة ونحسب لتكرارالها وتقوم بإجراء الحسابات اللازمة بعد الأخذ بالحسبان التفتيلات اللازمة.

ونتم عملية التقدير كما يلي: لتفترض أننا نريد تقدير معدل نصيب الفــرد مــن المساحة السكنية في مجتمع ما عدد أسرة N. فمن أجل ذلك نقــوم بســحب عينــة عشوائية بسيطة من هذه الأسر بحجم n ، ونأخذ المعلومات اللازمة عن كل أســرة ، ولنرمز للمساحة السكنية التي تشغلها الأسرة i بـــ y_i ولعدد أفراد تلك الأســرة بـــ m_i ، فإن معدل نصيب الفرد من المساحة السكنية في العينة المسحوبة يســـاوي مجموع المساحات السكنية على مجموع أفراد أسر العينة أي:

$$t=\frac{\sum_{i=1}^{n}y_i}{\sum_{i=1}^{n}m_i}=\frac{\overline{y}}{\overline{m}}$$

حيث $\overline{m}$ و $\overline{y}$ هما المتوسطان الحسابيان لـــ m_i و y_i في العينة ؛ وإذا رمزنـــا بـــ y_i للمساحة السكنية التي تشغلها الأسرة i في المجتمع و M_i لعدد أفراد تلـــك الأسرة في المجتمع أيضا ، فإن نصيب الفرد من المساحة السكنية في المجتمع يساوي

$$T=\frac{\sum_{i=1}^{N}y_i}{\sum_{i=1}^{N}m_i}=\frac{\overline{Y}}{\overline{M}}$$

حيث $\overline{M}$ و $\overline{Y}$ هما المتوسطان الحسابيان لـــ M_i و Y_i في المجتمع. فـإذا كان حجم العينة كبيرا وكان المجتمع متجانسا فإن المقدار t يعد تقديرا غير منحاز لـــ T ويمكننا أن نكتب:

$\hat{T}=t=\frac{\overline{y}}{\overline{m}}$ وأن هذه العلاقة هي التي تعطينا تقدير المعدل المطلوب.

أما تباين التقدير t فيمكن أن نحصل عليه كما يلي : لنأخذ

$$t-T=\frac{\overline{y}}{\overline{m}}-T=\frac{\overline{y}-\overline{m}T}{\overline{m}}$$

وإذا كان حجم العينة n كبيرا فإن $\overline{m}$ لا يمكن أن يختلف كثيرا عن $\overline{M}$ لذلـــك يمكننا أن نكتب:

$$t-T\approx\frac{\overline{y}-\overline{m}T}{\overline{M}}$$

حيث $$\overline{M}=\frac{\sum_{i=1}^{N}M_i}{N}$$

وبذلك نجد أن :

$$\sigma_t^2=E(t-T)^2\approx E\left(\frac{\overline{y}-\overline{m}T}{\overline{M}}\right)^2$$

$$=\frac{1}{\overline{M}^2}E(\overline{y}-\overline{m}T)^2$$

وإذا علمنا أن المقدار $\overline{y}-\overline{m}T$ مــــا هـــو إلا المتوســـط الحســـابي للمقـــدار $u_i=y_i-m_iT$ المأخوذ من العينة ، ومتوسط هذا المقدار في المجتمع يساوي:

$$\overline{U}=\frac{1}{N}\sum_{i=1}^{N}u_i=\frac{1}{N}\sum_{i=1}^{N}y_i-\frac{1}{N}T\sum_{i=1}^{N}m_i$$

$$\Rightarrow\quad \overline{U}\quad=\overline{Y}-T\overline{M}=0$$

وبأخذ $\overline{u}=\overline{y}-\overline{m}T$ فيمكننا أن نكتب عندئذ:

$$\sigma_t^2=\frac{1}{\overline{M}^2}E(\overline{u}-\overline{U})^2$$

وباعتماد حسابات مشابه في العلاقات في الفقرات السابقة يمكننا أن نجد في حالة السحب بدون إعادة أن:

$$\sigma_t^2=\frac{1}{\overline{M}^2}.\frac{N-n}{N}.\frac{S_u^2}{n}$$

حيث أن :

$$S_u^2=\frac{1}{N-1}\sum_{i=1}^{N}(u_i-\overline{U})^2$$

وبما أن $\overline{U}=0$ عندئذ

$$S_U^2=\frac{1}{N-1}\sum_{i=1}^{N}(y_i-m_i T)^2$$

ومن الدراسات السابقة نجد أن التقدير غير المنحاز والمتماسك لــ S_U^2 هو تباين العينة المعرف بالعلاقة.

$s_u^2=\frac{1}{n-1}\sum_{i=1}^{n}(y_i-m_i t)^2$ وهكذا نحصل على تقدير لتباين المعدل t في حـــالة السحب بدون إعادة من الشكل:

$$\hat{\sigma}_t^2=\frac{1}{\overline{M}^2}.\frac{N-n}{N}.\frac{s_u^2}{n}$$

وبالتحديد نجد:

$$\hat{\sigma}_t^2=\frac{1}{\overline{M}^2}.\frac{N-n}{N}.\frac{1}{n(n-1)}\sum_{i=1}^{n}(y_i-m_i t)^2$$

وفي حالة السحب مع الإعادة يكون تقدير التباين من الشكل

$$\hat{\sigma}_t^2=\frac{1}{\overline{M}^2}.\frac{s_u^2}{n}=\frac{1}{\overline{M}^2}.\frac{1}{n(n-1)}.\sum_{i=1}^{n}(y_i-m_i t)^2$$

أما فيما يتعلق بتقدير الانحراف المعياري فإنه يساوي في حالـــة الســـحب مـــع الإعادة:

$$\hat{\sigma}_t=\frac{1}{\overline{M}}\sqrt{\frac{1}{n(n-1)}}.\sqrt{\sum_{i=1}^{n}(y_i-m_i t)^2}$$

وفي حالة السحب بدون إعادة يكون من الشكل:

$$\hat{\sigma}_t=\frac{1}{\overline{M}}\sqrt{\frac{N-n}{Nn(n-1)}}.\sqrt{\sum_{i=1}^{n}(y_i-m_i t)^2}$$

ويكون مدى الثقة للتقدير t من الشكل:

$$T\in[t-Z\hat{\sigma}_t\ ,\ t+Z\hat{\sigma}_t]$$

أو $\quad t-Z\hat{\sigma}_t \leq T \leq t+Z\hat{\sigma}_t$

مثال:

سحبنا مع الإعادة عينة عشوائية بسيطة من الحجم 33 أسرة من سكان مدينة ما ، وأخذنا المعلومات اللازمة عن عدد أفراد أسرها m_i وعن مقدار دخلها الشـهري y_{1i} وعن مقدار انفاقها الشهري على الأغذية y_{2i} ورتبنا هذه المعلومات في الجدول التالي:

رقم الأسرة i	عدد أفراد الأسرة m_i	الدخل الشهري للأسرة y_{1i}	الإنفاق الشهري للأسرة على الأغذية y_{2i}
1	2	62	14.3
2	3	62	20.8
3	3	87	22.7
4	5	55	30.5
5	4	58	41.2
6	7	92	28.2
7	2	88	24.2
8	4	79	30.0
9	2	83	24.2
10	5	62	14.4
11	3	63	13.4
12	6	62	19.8
13	4	60	29.4
14	4	75	27.1
15	2	90	22.2
16	5	75	37.7
17	3	69	22.6
18	4	83	63.0
19	2	85	20.6
20	4	73	27.7
21	2	66	25.9
22	5	58	23.3
23	3	77	39.8
24	4	69	16.8

25	7	65	37.8
26	3	77	34.8
27	3	69	28.7
28	6	95	63.6
29	2	77	19.6
30	2	69	21.6
31	6	69	18.2
32	4	67	20.1
33	2	63	20.7
المجموع	123	2394	9.07.2

والمطلوب:

1- حساب تقدير متوسط إنفاق الأسرة على الأغذية.

2- حساب تقدير معدل نصيب الفرد من الإنفاق على الأغذية.

3- حساب النسبة المئوية للإنفاق على الأغذية من الدخل.

4- حساب تقدير الانحراف المعياري لكل من التقديرات السابقة.

5- حساب تقدير ثابت التباين وتقدير مدى الثقة باحتمال قـــدره 0.95 لمعـــدل نصيب الفرد من للإنفاق.

الحل:

1- إن تقدير متوسط إنفاق الأسرة على الأغذية ما هو إلا المتوســـط الحســـابي لإنفاق في العينة حيث نجد أن

$$\bar{y}_2 = \frac{\sum_{i=1}^{33} y_{2i}}{33} = 27.49$$

ولتقدير الانحراف المعياري لهذا المتوسط نحسب من:

$$\sigma_{\bar{y}_2}=\frac{s}{\sqrt{n}}=\frac{1}{\sqrt{n}}\sqrt{\frac{\sum_{i=1}^{n}(y_{2i}-\bar{y}_2)^2}{n-1}}$$

$$=\frac{1}{\sqrt{n(n-1)}}.\sqrt{\sum_{i=1}^{n}y_{2i}^2-\frac{\left(\sum_{i=1}^{n}y_{2i}\right)^2}{n}}$$

$$=\frac{1}{\sqrt{(33)(32)}}.\sqrt{28224-\frac{(907.2)^2}{33}}=1.76$$

2- لتقدير معدل نصيب الفرد من الإنفاق على الأغذية فإننا نجد أن:

$$t=\frac{\sum_{i=1}^{n}y_{2i}}{\sum_{i=1}^{n}m_i}=\frac{907.2}{123}=7.38$$

ولتقدير الانحراف المعياري لهذا المعدل نستخدم العلاقة:

$$\hat{\sigma}_t=\frac{1}{\overline{M}}.\frac{1}{\sqrt{n(n-1)}}.\sqrt{\sum_{i=1}^{n}(y_{2i}-tm_i)^2}$$

$$=\frac{1}{\overline{M}.\sqrt{n(n-1)}}.\sqrt{\sum_{i=1}^{n}y_{2i}^2-2t\sum_{i=1}^{n}y_{2i}\,m_i+t^2\sum_{i=1}^{n}m_i^2}$$

وبما أن:

$$\sum_{i=1}^{n}m_i^2=533 \qquad \overline{M}\approx\overline{m}=3.7273$$

$$\sum_{i=1}^{n}y_{2i}^2=28224 \;\; ; \;\; \sum_{i=1}^{n}y_{2i}.m_i=3595.5$$

عندئذ:

$$\hat{\sigma}_t=\frac{1}{(3.7273)\sqrt{(33)(32)}}.\sqrt{28224-2(7.38)(3595.5)+(7.38)^2\,(533)}$$

$$=0.534$$

ومدى الثقة المطلوب

$$I_t = [7.38-(0.534)(2) \ , \ (7.38)+(0.534)(2)]$$
$$\Rightarrow T \in [6.312 \ , \ 8.448]$$

أي نكون واثقين 95% من أن معدل نصيب الفرد من الإنفاق لن يقل عـــن 6.312 ولن يزيد على 8.448 (وحدة نقدية). وثابت التباين يساوي

$$C = \frac{\hat{\sigma}_t}{t} = \frac{0.534}{7.38} = 0.072$$

وهذا ما يدل على فعالية التقدير t لأن ثابت التباين أصغر من 0.1.

3- أما بالنسبة لتقدير النسبة المئوية للإنفاق فنجدها من:

$$P = \frac{\sum_{i=1}^{n} y_{2i}}{\sum_{i=1}^{n} y_{1i}} = \frac{907.2}{2394} = 0.379$$

والانحراف المعياري لهذه النسبة يحسب من

$$\hat{\sigma}_P = \frac{1}{\bar{y}_1\sqrt{n(n-1)}} \cdot \sqrt{\sum_{i=1}^{n} y_{2i}^2 - 2P\sum_{i=1}^{n} y_{1i}y_{2i} + P^2\sum_{i=1}^{n} y_{1i}^2}$$

حيث

$$\bar{y}_1 = 72.545 \ ; \ \sum_{i=1}^{n} y_{1i}.y_{2i} = 66678$$
$$\sum_{i=1}^{n} y_{1i}^2 = 177254 \ ; \ P = 0.379 \ ; \ \sum_{i=1}^{n} y_{2i}^2 = 28224$$

وبالتالي:

$$\hat{\sigma}_P = \frac{1}{(72.545)\sqrt{(33)(32)}} \cdot \sqrt{28224 - 2(0.379)(66678) + (0.379)^2 (177254)}$$
$$= 0.0238$$

14-2 : دراسة ارتباط خاصتين أو أكثر:

بعد الإطلاع على عملية الحصول على تقدير قيم خاصة معينة من جميع واحدات العينة ، نهتم الآن بالبحث عن ارتباط خاصتين معينتين في واحدات العينة أو أكثر، فمثلا عندما تجرى معاينة على أسر مدينة ما، فإننا نهتم بالحصول على عدد أفراد الأسر المدروسة ودخلها وكمية انفاقها على الأغذية والمساحة السكنية التي تشغلها. وكذلك الأمر عندما نقوم بدراسة واحدات بضاعة ما فإننا نهتم مثلا بوزنها وكلفتها وسعرها ونسبة العطب في تلك الواحدات. وهنا ندرس مثلا علاقة كمية الإنفاق بالدخل أو علاقة سعر البضاعة بكلفتها أو بمقدار الطلب عليها... إلخ. ولكننا نعلم من نظرية الارتباط أن البحث عن العلاقة الارتباطية يجب أن يكون موضوعيا وليس مجردا ، بمعنى أن قيم أحد الخواص يجب أن تكون مؤشرا حقيقيا في تحديد قيم الخاصة الآخرى. فإذن علينا قبل أن نبحث عن العلاقة الارتباطية أن نناقش بموضوعية كاملة ونبحث عن الخاصة المؤشرة والخاصة المتأثرة أي يجب أن نحدد المؤشر المسبب والمؤثر الناتج ، ثم نعد المؤشر الأول (المسبب) هو المتغير المستقل والثاني (الناتج) هو المتغير التابع.

وما يعترضنا في العمل هو السؤال التالي: هل تعد المعادلة الرياضية التي تمثل العلاقة الارتباطية بين قيم الخواص المأخوذة من العينة تقديرا غير متحيز للمعادلة الرياضية العامة التي تمثل العلاقة الارتباطية نفسها بين قيم تلك الخواص ولكن مأخوذة على جميع واحدات المجتمع.

وللإجابة على ذلك السؤال سنعرض بادىء الأمر بعض التعاريف الأساسية:

لنرمز بـ y_{1i} لقيم خاصة. الأولى والمأخوذة من واحدات العينة.

y_{2i} لقيم الخاصة لثانية والمأخوذة من واحدات العينة

ونعلم أن $\bar{y}_1 = \dfrac{\sum_{i=1}^{n} y_{1i}}{n} \quad ; \quad \bar{y}_2 = \dfrac{\sum_{i=1}^{n} y_{2i}}{n}$

وأن $E(\bar{y}_1) = \bar{Y}_1 \quad ; \quad E(\bar{y}_2) = \bar{Y}_2$

حيث إن $\overline{Y}_1$, $\overline{Y}_2$ هما المتوسطان الحسابيان لقيم الخاصة الأولى والثانية على الترتيب والمأخوذتان على جميع واحدات المجتمع الكلي.

وللبحث عن ارتباط الخاصتين Y_1 , Y_2 في المجتمع نبدأ بحساب قيمة معامل الارتباط لهما. ونحن نعلم من نظرية الارتباط أن معامل الارتباط البسيط بين الخاصتين Y_1 , Y_2 في المجتمع يعرف بالعلاقة:

$$R_{Y_1,Y_2} = \frac{\sum_{i=1}^{N}(Y_{1i} - \overline{Y}_1)(Y_{2i} - \overline{Y}_2)}{N\sigma_{Y_1}.\sigma_{Y_2}}$$

ولنبحث عن تقدير غير منحاز لـ R_{Y_1,Y_2} من معطيات العينة ، لذلك نبحث عن التقديرات غير المنحازه لكل من الرموز الداخلة في هذه العلاقة . حيث نجد أن: تباين قيم خاصة ما في العينة تعرف بالعلاقة

$$s^2 = \frac{\sum_{i=1}^{n}(y_i - \overline{y})^2}{n-1}$$

وهو تقدير غير منحاز لتباين تلك الخاصة في المجتمع σ_Y^2 أما بالنسبة للجداء:

$$\frac{\sum_{i=1}^{N}(y_{1i} - \overline{Y}_1)(y_{2i} - \overline{Y}_2)}{N}$$

والذي يدعى بتمام التبياين فإن تقديره غير المنحاز يعطى في نظرية العينات بالعلاقة التالية:

$$CoV_{(y_1,y_2)} = \frac{\sum_{i=1}^{n}(y_{1i} - \overline{y}_1)(y_{2i} - \overline{y}_2)}{n-1}$$

ومما سبق نستنتج أنه يمكننا أن نقدر معامل الارتباط في نظرية العينات بالعلاقة التالية:

$$\hat{R}_{(y_1,y_2)}=\frac{\sum_{i=1}^{n}(y_{1i}-\bar{y}_1)(y_{2i}-\bar{y}_2)}{(n-1)s_{y_1}.s_{y_2}}$$

والتي من الممكن كتابتها كما يلي:

$$\hat{R}_{(y_1,y_2)}=\frac{\sum_{i=1}^{n}(y_{1i}-\bar{y}_1)(y_{2i}-\bar{y}_2)}{\sqrt{\sum_{i=1}^{n}(y_{1i}-\bar{y}_1)^2.\sum_{i=1}^{n}(y_{2i}-\bar{y}_2)^2}}$$

وهذه العلاقة لا تختلف عن العلاقة المعروفة والتي تعرف معامل الارتباط البسيط في حالة الإحصاء الشامل.

ولإيجاد المعادلة الرياضية التي تمثل العلاقة الإرتباطية بين قيم y_1 , y_2 نبحث عن المؤشر المسبب وليكن y_1 ومن معلومات العينة نرسم شكل الانتشار لقيم هاتين عن المعادلة التي يجب أن نبحث عنها لتمثيل هذه العلاقة الارتباطية والتي قد تكون من أحد الأشكال التالية:

$$\hat{y}_{2i}=\alpha+\beta y_{1i}$$

$$\hat{y}_{2i}=\alpha+\beta y_{1i}+6y_{1i}^2$$

$$\hat{y}_{2i}=\alpha\, e^{\beta y_{1i}}$$

$$\hat{y}_{2i}=\alpha\, y_{1i}^{\beta}$$

$$\hat{y}_{2i}=\alpha+A\sin y_{1i}+\beta_{ws}y_{1i}$$

حيث نرمز $\hat{y}_{2i}$ لقيمة y_{2i} النظرية المستخرجة من المعادلة (أي المقدرة من المعادلة). لنفترض أن شكل الانتشار المرسوم من معلومات العينة سمح لنا بأن نفترض أن علاقة قيم Y_1 , Y_2 في المجتمع هي من الشكل: $Y_{2i}=\alpha+\beta Y_{1i}$

والمطلوب عندئذ تقدير كل من α , β من معطيات العينة. ولإيجاد مثل هذه التقديرات نعود ونفترض أن علاقة قيم y_2 , y_1 هي خطية أيضا من الشكل:

$$y_{2i}=\hat{\alpha}+\hat{\beta}y_{1i}$$

والمطلوب عندئذ البحث عن قيمة لكل من $\hat{\alpha}$, $\hat{\beta}$ بحيث تكون هاتان القيمتان تمثلان تقديرين غير منحازين لكل من α , β على الترتيب ، وإيجاد تلـك القيـم بوساطة طريقة المربعات الصغرى والتي تعطينا المعادلات الشهيرة التالية:

$$\sum_{i=1}^{n} y_{2i} = n\hat{\alpha} + \hat{\beta} \sum_{i=1}^{n} y_{1i}$$

$$\sum_{i=1}^{n} y_{1i}\, y_{2i} = \hat{\alpha} \sum_{i=1}^{n} y_{1i} + \hat{\beta} \sum_{i=1}^{n} y_{1i}^{2}$$

وبحل هاتين المعادلتين نحصل على قيمة لـ $\hat{\alpha}$, $\hat{\beta}$ تعد كل منهما تقديرا غـير منحاز ومتماسك وفعال أيضا لـ α , β على الترتيب. وذلك لأن طريقة المربعـات الصغرى تعتمد أساسا على إيجاد قيم الثوابت التي تحقق أصغر تباين ممكـن وبذلـك يمكننا أن نمثل العلاقة الإرتباطية بين Y_1 , Y_2 بالعلاقة :

$$\hat{y}_{2i} = \hat{\alpha} + \hat{\beta}\ \hat{y}_{1i}$$

وواضح أن هذه العلاقة هي تقدير غير منحاز لـ Y_{2i}.

ولكن هذا التمثيل يتضمن أخطار عديدة فبالإضافة إلى الأخطاء الناجمـة عـن السحب والقياس والحسابات هناك أخطاء أخرى ناجمة عن إدخال كل من التقديريـن $\hat{\alpha}$, $\hat{\beta}$ وهذه الأخطاء يمكن تقديرهما إذا علمنا تقدير تباين كل مـن $\hat{\alpha}$, $\hat{\beta}$. غير أن حساب أخطاء كل من $\hat{\alpha}$, $\hat{\beta}$ وبشكل عام قد لا يفيدنا كثيرا في البحوث التطبيقية، حيث المراد هو حساب مجمل الخطأ المرتكب في تقدير Y_2 بفرض متغـير عشوائي نرمز له بـ u_i نضيفه إلى معادلة التمثيل حيث نحصل على المعادلة التالية:

$$Y_{2i} = \hat{Y}_{2i} + u_i = \hat{\alpha} + \beta y_{1i} + u_i$$

وبذلك $u_i = Y_{2i} - \hat{Y}_{2i}$ وتباينه يساوي

$$\sigma_u^2 = \frac{\sum_{i=1}^{N} (Y_{2i} - \hat{Y}_{2i})^2}{N}$$

وإن تقدير هذا التباين يعطي عندئذ بالعلاقة

$$\hat{\sigma}_u^2=\frac{\sum_{i=1}^{n}(y_{2i}-\hat{y}_{2i})^2}{n-2}$$

وهنا يمكننا عند التنبؤ بقيمة ما لـ Y_2 المقابلة لقيمة ما معطاة لـ Y_1 أن تعد وبخطأ مقبول ضمن المجال

$$\left[\hat{Y}-Z\hat{\sigma}_u\,,\ \hat{Y}+Z\hat{\sigma}_u\right]$$

وهذا يمثل مجال الثقة الموافق للمعامل Z هنا تحدد وفقا للاحتمال المطلوب.

ولكن إذا أردنا التنبؤ بقيمة Y_2 المقابلة لقيمة معلومة y_{1P} لـ Y_1 (أي عندما $Y_1=y_{1P}$) فإن تقدير σ_u^2 يعطي بالعلاقة التالية:

$$\hat{\sigma}_u^2=\left[\frac{\sum_{i=1}^{n}(y_{2i}-\bar{y}_{2i})^2}{n-2}\right].\left[1+\frac{1}{n}+\frac{(y_{1P}-\bar{y}_1)^2}{\sum_{i=1}^{N}(y_{1i}-\bar{y}_1)^2}\right]$$

ومن الملاحظ أن العلاقة الأخيرة تتأثر بقيمة y_{1P} والتي يتم التنبؤ عندها.

مثال:

لنفترض أننا سحبنا عينة مؤلفة من سبع أسر في مدينة ما وسجلنا دخل وإنفاق كل أسرة وحصلنا على النتائج التالية:

رقم الأسرة	1	2	3	4	5	6	7
الدخل بالوحدة النقدية	200	250	270	300	320	350	370
الإنفاق بالوحدة النقدية	150	170	200	250	260	280	300

والمطلوب:

1- أوجد العلاقة الارتباطية بين الدخل والإنفاق.

2- عين حدود الثقة للعلاقة الرياضية المصممة على المجتمع ككل.

الحل:

لتمثيل هذه العلاقة الارتباطية للدخل y_{1i} وللنفقات y_{2i} نفترض أن معادلة التمثيل من الشكل

$$\hat{y}_{2i} = \hat{\alpha} + \hat{\beta} y_{1i}$$

ثم نحسب الثوابت $\hat{\beta}$, $\hat{\alpha}$ من المعادلات العادية التالية:

$$\sum_{i=1}^{n} y_{2i} = n\hat{\alpha} + \hat{\beta} \sum_{i=1}^{n} y_{1i}$$

$$\sum_{i=1}^{n} y_{1i}\ y_{2i} = \hat{\alpha} \sum_{i=1}^{n} y_{1i} + \hat{\beta} \sum_{i=1}^{N} y_{1i}$$

ولأجل ذلك تنظم جدول الحسابات التالي:

رقم الأسرة	y_{1i}	y_{2i}	$y_{1i} . y_{2i}$	y_{1i}^2	$\hat{y}_{2i}$	$y_{2i} - \hat{y}$	$(y_{2i} - \hat{y}_{2i})^2$
1	200	150	30000	40000	140.75	9.25	85.56
2	250	170	42500	62500	188.25	- 18.25	333.6
3	270	200	54000	72900	207.25	- 7.25	52.57
4	300	250	75000	90000	235.75	14.25	203.6
5	320	260	83200	102400	254.75	5.25	27.56
6	350	280	98000	122500	283.25	- 3.25	10.56
7	370	300	111000	136900	302.25	- 2.25	5.06
المجموع	2060	1610	493700	627200			717.43

وبالتالي :

$$\begin{rcases} 1610 = 7\hat{\alpha} + 2060\,\hat{\beta} \\ 492700 = 2060\ \hat{a} + 627200\ \hat{\beta} \end{rcases}$$

وبحل هاتين المعادلتين نجد أن:

$$\hat{\alpha} = -45.25 \quad ; \quad \hat{\beta} = 0.95$$

والمعادلة تصبح من الشكل :

$$\hat{y}_{2i} = -49.25 + 0.95 y_{1i}$$

ومن الجدول نجد أن:

$$\hat{\sigma}_u = \sqrt{\frac{717.43}{5}} \approx 12$$

ومما سبق نستنتج أنه إذا وضعنا علامة عامة لتمثيل ارتباط y_2 بـــ y_1 علـــى الشكل:

$$\hat{y}_{2i} = -49.25 + 0.95 y_{1i}$$

وأردنا التنبؤ بقيمة y_2 (الإنفاق) عندما نصـــادف مثـــلا أســـرة ذات دخـــل $y_1 = 400$ فإننا نجد أن:

$$\hat{y}_2 = -49.25 + 0.95(400) = 330.75$$

وإن مجال الثقة الـــ 95% يساوي:

$$\hat{y}_2 - Z\hat{\sigma}_u \leq \hat{Y}_2 \leq \hat{y}_2 + Z\hat{\sigma}_u$$

$$330.75 - 2(12) \leq \hat{Y}_2 \leq 330.75 + (2)(12)$$

$$306.73 \leq \hat{Y}_2 \leq 354.75$$

ومن نكون واثقين 95% من أجل أسره دخلها 400 وحدة نقدية لن يقل إنفاقـــها عن 306.73 ولن يزيد على 354.75.

15-2: تقدير الفرق بين متوسطي مجتمعين:

نفترض أنه لدينا مجتمعان إحصائيان منفصلان متوسط الأول $\bar{Y}_1$ ومتوسط الثـــاني $\bar{Y}_2$. فلتقدير هذين المتوسطين ومن ثم مقارنتهما يعمد الباحثون إلى سحب عينة مـــن كل منهما، فإذا فرضنا أن حجم العينة المسحوبة من المجتمع الأول هو n_1 ومتوســـط قيم الخاصة المدروسة فيها هو $\bar{y}_1$ وإن حجم العينة المسحوبة من المجتمع الثاني هو n_2 ومتوسط قيم الخاصة المدروسة نفسها هو $\bar{y}_2$. وحسب ما سبق إن $\bar{y}_1$ يعد تقديـــرا غير منحاز لـــ $\bar{Y}_1$ وإن $\bar{y}_2$ يعد تقديرا غير منحاز لـــ $\bar{Y}_2$ كما أن الفرق $(\bar{y}_1 - \bar{y}_2)$ يعد تقديرا غير منحاز للفرق $(\bar{Y}_1 - \bar{Y}_2)$ لأن:

$$E(\bar{y}_1 - \bar{y}_2) = E(\bar{y}_1) - E(\bar{y}_2) = \bar{Y}_1 - \bar{Y}_2$$

وبما أن العينتين المسحوبتين مستقلتان عندئذ:

$$\sigma^2_{(\bar{y}_1-\bar{y}_2)}=\sigma^2_{\bar{y}_1}+\sigma^2_{\bar{y}_2}$$

ونميز هنا حالتين:

1- حالة السحب بدون إعادة يكون:

$$\sigma^2_{(\bar{y}_1-\bar{y}_2)}=\frac{N_1-n_1}{N_1-1}.\frac{\sigma_1^2}{n_1}+\frac{N_2-n_2}{N_2-1}.\frac{\sigma_2^2}{n_2}$$

والذي يعطي تقديره بالعلاقة:

$$\hat{\sigma}_{(\bar{y}_1-\bar{y}_2)}=\frac{N_1-n_1}{N_1}.\frac{s_1^2}{n_1}+\frac{N_2-n_2}{N_2}.\frac{s_2^2}{n_2}$$

2- حالة السحب مع الإعادة : نجد:

$$\sigma^2_{(\bar{y}_1-\bar{y}_2)}=.\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}$$

وتقديره يعطي بالعلاقة:

$$\hat{\sigma}^2_{(\bar{y}_1-\bar{y}_2)}=\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}$$

حيث σ_1^2 , σ_2^2 هما تباينا المجتمعين و s_1^2 , s_2^2 هما تباينا العينتين المسحوبتين.

وتستخدم هذه العلاقات في تحديد مدى الثقة للفرق $(\bar{Y}_1-\bar{Y}_2)$ حسب ماورد سابقا

$$(\bar{y}_1-\bar{y}_2)Z\,\hat{\sigma}_{(\bar{y}_1-\bar{y}_2)}\leq\bar{Y}_1-\bar{Y}_2\leq(\bar{y}_1-\bar{y}_2)+Z\,\hat{\sigma}_{(\bar{y}_1-\bar{y}_2)}$$

وللمقارنة تميز ثلاث حالات

1- طرفا المجال موجبان $\Leftarrow \bar{Y}_1-\bar{Y}_2>0 \Leftarrow \bar{Y}_1>\bar{Y}_2$

2- طرفا المجال سالبان $\Leftarrow \bar{Y}_1-\bar{Y}_2<0 \Leftarrow \bar{Y}_1>\bar{Y}_2$

3- طرف موجب وطرف سالب $\Leftarrow$ وبثقة محددة يكون $\bar{Y}_1-\bar{Y}_2=0$

أي $\bar{Y}_1=\bar{Y}_2$

ومن هذه المقارنة نحدد النتائج المرجوة.

16-2: تمارين غير محلولة:

1- مجتمع مؤلف من العناصر التالية: 6, 3, 4, 1,1

وسحبت منه عينة عناصرها 1, 1, 4

احسب قيمة كل من s^2 , $\sigma^2_{\bar{y}}$, $\hat{\sigma}^{\varepsilon}_{\bar{Y}}$.

2- لنفترض أننا سحبنا عينة من الأسر لتقدير وسطي مصروفهم الشهري على الأغذية واشترط علينا أن تكون الدقة d= 4 ل.س ودرجة من الثقة قدرها 0.99 ، فما هو حجم العينة المطلوب سحبها لتلبي الشروط السابقة علما بأن S= 12 وحجم المجتمع N= 2000 وأن السحب يجري بدون إعادة.

ثم عين حجم العينة في حالة السحب مع الإعادة.

3- سحبنا من مجتمع N= 50 عينة ذات حجم n= 10 بدون إعادة وكانت النتائج كما يلي : 7 ,10 ,9 ,5 ,11 ,7 ,8 ,4 ,10 ,12.

احسب قيمة $\bar{y}$ وقدر قيمة Y واحسب s^2 ثم احسب $\hat{\sigma}^2_Y$ واحسب مدى الثقة لـ Y في حدود $3\hat{\sigma}_{\hat{Y}}$.

4- نفترض أننا نجمع تفاحات بستان ما في سلال وقدرها 4000 سلة وأظهرت نتائج عينة عشوائية سحبت منها 50 سلة أن وسطي عدد التفاحات في كل سلة هو $\bar{y}$=44 وأن S= 5 تفاحات علما بأن السحب يجري بدون إعادة. والمطلوب:

1- قدر قيمة Y وقيمة $\hat{\sigma}_{\hat{Y}}$.

2- احسب حجم العينة الضرورية حتى لا يتجاوز الانحراف المعياري لـ Y مقدار 2000 تفاحة وذلك باحتمال 95% (أي Z= 2).

5- من مجتمع عدد عناصره 2000 فرد سحبت عينة بدون إعادة قدرها 200 فرد واستقصيت آرائهم حول أحد الاقتراحات فكان 120 منهم مؤيدين لذلك الاقتراح 57 منهم ضده و23 لم يدلوا بآرائهم. والمطلوب إيجاد تقدير لنسبة مؤيدي ذلك الاقتراح في المجتمع المدروس وحساب مدى الثقة،لهذا التقدير الذي يحقق لنا ثقة باحتمال قدره 0/95. ثم عين مدى الثقة لنسبة الأفراد الذين يعارضون ذلك الإقتراح والذي يحقق لنا ثقة باحتمال قدره 0.99 مع العلم بأن الذين لم يدلوا بأصواتهم يمكن عدهم من

المعارضين ثم احسب حجم العينة الذي كان يجب سحبها لتحقيق الدقـــة الســابقة وبالاحتمال السابقة.

6- مجتمع مؤلف من 4500 عنصر وتباين يساوي 100. عين حجم العينة الـــلازم سحبها لتقدير متوسط المجتمع بدقة d= 5 واحتمال 0.95 ثم عين حجم العينة الـــلازم لتقدير إجمالي المجتمع بالدقة والاحتمال نفسيهما.

7- سحبت عينة عشوائية بسيطة حجمها 5 عناصر من مجتمع إحصائي يتألف مـن 100 عنصر. وكانت قيم الخاصة المدروسة في هذه العينة: 10, 8, 6, 4, 2 والمطلوب:

1- قدر متوسط المجتمع والقيمة الإجمالية لوحدات المجتمع.

2- احسب تباين العينة وانحرافها المعياري.

3- احسب الخطأ المعياري لتقدير متوسط المجتمع.

4- احسب الانحراف المعياري للقيمة الكلية لوحدات المجتمع.

8- مجتمع إحصائي بحجم N= 6 وقيم عناصره:

7, 4, 11, 1, 3, 8 احسب:

1- متوسط العينة $\bar{y}$ في كل من العينات العشوائية البسيطة الممكنة والتي حجمـها n= 2.

2- تحقق من أن $\bar{y}$ هو تقدير غير منحاز لـ $\bar{Y}$ واحسب تباين $\bar{y}$.

3- احسب s^2 لكل من العينات العشوائية البسيطة الممكنة والتي حجمها 3 وتحقق من أن $E(s^2)=S^2$

4- إذا كانت العينات العشوائية ذات الحجم 2 مسحوبة مع الإعادة مـــن هـــذا المجتمع، فبين عن طريق إيجاد كل العينات الممكنة أن $V(\bar{y})$ يحقق المعادلة

$$V(\bar{y})=\frac{\sigma^2}{n}=\frac{s^2}{n}\left(\frac{N-1}{N}\right)$$

9- من قائمة من 468 من الكليات الصغيرة ذات السنتين، سحبنا عينة عشـــوائية بسيطة من 100 كلية ونضمت هذه العينة 54 كلية عامة و 46 كلية خاصة. والبيـــان المتعلق بعدد الطلاب y وعدد المدرسين n هو كما يلي:

	n	$\sum y$	$\sum n$	$\sum y^2$	$\sum yx$	$\sum x^2$
عام	54	31281	2024	29881219	1729349	111090
خاص	46	13707	1075	6366785	431041	33119

والمطلوب:

1- في كل نوع من الكليات قدر نسبة عدد الطلاب إلى عدد الأساتذة.

2- احسب الأخطاء المعيارية لتقديراتك.

3- في الكليات العامة، عين 90% حدود ثقة لنسبة الطلاب إلى المدرسين في المجتمع بكامله.

4- اختبر عند المستوى 5% ما إذا كانت نسبة الطلاب إلى المدرسين مختلفة بصورة مهمة في النوعين من الكليات.

5- في الكليات العامة، قدر العدد الكلي للمدرسين:

a- علما بأن العدد الكلي للكليات العامة في المجتمع هو 251.

b- بدون معرفة هذا الرقم وفي كل حالة احسب الخطأ المعياري لتقديرك.

10- سنقوم بمسح إحصائي صغير لمقارنة مالكي المنازل مع المستأجرين. ويوجد في المجتمع نحو 75% من المالكين و25% من المستأجرين. ومن أجل أحد مفردات البحث يعتقد بأن التباين هو نحو 15 لكل من المالكين والمستأجرين والخطأ المعياري للفرق بين متوسطين الميدانين يجب أن يتجاوز الواحد. عين حجم العينة التي نحتاجها في الحالات التالية:

1- إذا أمكن تحديد المالكين والمستأجرين قبل سحب العينة.

2- إذا لم يكن هذا ممكنا.

11- من قائمة تحوي 3042 اسما ، تم سحب عينة عشوائية بسيطة (بدون إعادة) من الحجم 200 وتبين أن 38 عنوانا خاطئا. والمطلوب:

1- عين مقدار العدد الكلي للعناوين الموجودة في القائمة والتي تحتاج إلى تصحيح.

2- عين الخطأ المعياري لهذا المقدر

ملاحظة: (نستخدم هنا المعلومات التالية):

$$n=200;\ N=3042;\ a=38;\ r=\frac{a}{n}$$

$$\hat{A}=Nr$$

$$\left(S_{\hat{A}}=\sqrt{\frac{N(N-n)}{n-1}.r.q}\right.$$

الفصل الثالث
المعاينة العشوائية الطبقية

3-1: مقدمة:

أول ما نقوم به في المعاينة الطبقية هو تقسيم المجتمع المؤلف من N عنصراً إلى مجتمعات جزئية فيها $N_1, N_2, ..., N_L$ من العناصر على الترتيب وهذه المجتمعات الجزئية غير متداخلة وهي تؤلف مع بعضها المجتمع بكامله.

أي أن $N=\sum_{h=1}^{L} N_h$. وتدعى المجتمعات الجزئية طبقات ، وللحصول على الفائدة التامة من عملية التقسيم إلى طبقات يجب معرفة قيم المقادير N_h ، وعند تحديد الطبقات تسحب عينة من كل طبقة، ويتم السحب بصورة مستقلة في الطبقات المختلفة. ونرمز لحجموم العينات ضمن الطبقات بـ: $n_1, n_2, ..., n_L$ على الترتيب. وإذا أخذنا عينة عشوائية بسيطة من كل طبقة، نصف عندئذ مجمل الطريقة بأنها معاينة عشوائية طبقية.

والتقسيم إلى طبقات طريقة عامة جدا ، وهناك أسباب كثيرة لذلك أهمها مايلي:

a- إذا أردنا معلومات إحصائية، وبدقة معروفة لأجزاء معينة من المجتمع، فمن المستحسن أن نعالج كل جزء وكأنه مجتمع قائم بذاته.

b- وقد تملي الراحة في العمل الإداري استخدام التقسيم إلى طبقات فمثلا قد يكون للوكالة التي تقوم بمسح إحصائي دوائر ميدانية نشرف كل دائرة منها على المسح المتعلق بجزء من المجتمع.

c- قد تختلف مشكلات المعاينة بصورة ملحوظة في أجزاء مختلفة من المجتمع. وفي المجتمعات البشرية، غالبا ما يوضع الأشخاص الذين يعيشون في مؤسسات (فنادق — مشافي — سجون) في طبقة مختلفة عن أولئك الذين يعيشون في بيوت عادية ، لأن طرائق المعاينة المناسبة للحالتين مختلفة.

d- يمكن أن يؤدي التقسيم إلى طبقات إلى كسب في دقة تقديرات صفات مميّزة للمجتمعات ككل، والفكرة الأساسية هي أنه قد يكون من الممكن تقسيم مجتمع غير متجانس إلى مجتمعات جزئية متجانسة داخلياً وهذا ما يوجب به اسم الطبقات وتعالج نظرية المعاينة الطبقية خواص التقديرات من عينة طبقية وأفضل اختيار لحجوم العينات بحيث نحصل على أعظم دقه ممكنة.

2-3: بعض الرموز المستخدمة في المعاينة الطبقية:

يرمز للدليل h للطبقة ولـ i للعنصر ضمن الطبقة والرموز هي تعميم طبيعي لتلك المستخدمة في الفصول السابقة. وتشير جميع الرموز التالية إلى طبقة h:

N_h تعني العدد الكلي للعناصر.

n_h عدد عناصر العينة المسحوبة من الطبقة h.

y_{hi} القيمة التي تحصل عليها من أجل العنصر i في العينة من الطبقة h.

$W_h = \frac{N_n}{N}$ ترجيحة الطبقة.

$f_h = \frac{n_h}{N_h}$ كسر المعاينة ضمن الطبقة.

$\overline{Y}_h = \frac{\sum_{i=1}^{N_h} y_{hi}}{N_h}$ المتوسط الصحيح.

$\overline{y}_h = \frac{\sum_{i=1}^{n_h} y_{hi}}{n_h}$ متوسط العينة

$S_h^2 = \frac{\sum_{i=1}^{N_h} (y_{hi} - \overline{Y}_h)^2}{N_h - 1}$ التباين الصحيح.

3-3: التقديرات وخواصها في المعاينة الطبقية:

من أجل متوسط المجتمع، نجد أن التقدير المستخدم في المعاينة الطبقية هو $\overline{y}_{st}$ حيث:

$$\cdot \bar{y}_{st} = \frac{\sum_{h=1}^{L} N_h \bar{y}_h}{N} = \sum_{h=1}^{L} W_h \cdot \bar{y}_h$$

وحيث $\left(\sum_{h=1}^{L} N_h = N\right)$

ولا يكون التقدير $\bar{y}_{st}$ بصورة عامة، هو متوسط العينة نفسها وذلك لأنه يمكـــن كتابة متوسط العينة $\bar{y}$ على الشكل:

$$\bar{y} = \frac{\sum_{h=1}^{L} n_h \bar{y}_h}{n}$$

والفرق هو أنه في $\bar{y}_{st}$ نتلقى التقديرات من الطبقات كل بمفردهـــا ترجيحاتهـــا الصحيحة $\frac{N_h}{N}$. ومن الواضح أنه تتطابق مع $\bar{y}_{st}$ في حالة كون $\frac{n_h}{n} = \frac{N_h}{N}$ في كل طبقة أو $\frac{n_h}{N_h} = \frac{n}{N}$ أو $f_h = f$

وهذا يعني أن كسر المعاينة يبقى نفسه في جميع الطبقات. ويوصف التقســـيم إلى طبقات كهذا بأنه تقسيم إلى طبقات بحصص متناسبة مع n_h أو المحاصة التناســـبية. وهذا يعطي عينة ذاتية الترجيح. وإذا كان لدينا العديد من التقديرات لنقوم بها ، فـــإن العينة ذاتية الترجيح توفر الوقت.

ونبيّن في المبرهنات التالية الخواص الرئيسة للتقدير $\bar{y}_{st}$.

مبرهنة (3-1):

إذا كان تقدير العينة $\bar{y}_h$ غير منحاز في كل طبقة، سيكون $\bar{y}_{st}$ تقديراً غير منحاز لمتوسط المجتمع $\bar{Y}$.

الإثبات:

$$E(\bar{y}_{st}) = E\left(\sum_{h=1}^{L} W_h \, \bar{y}_n\right) = \sum_{h=1}^{L} W_h \, \bar{Y}_h$$

بفرض أن التقديرات في كل طبقة هي تقديرات غير منحازة. ولكن يمكن كتابـــة متوسط المجتمع $\bar{Y}$ بالشكل التالية:

$$\overline{Y}=\frac{\sum_{h=1}^{L}\sum_{i=1}^{N_h}y_{hi}}{N}=\frac{\sum_{h=1}^{L}N_h\overline{Y}_h}{N}=\sum_{h=1}^{L}W_h\overline{Y}_h$$

ومن نجد أن: $E(\overline{y}_{st})=\overline{Y}$

مبرهنة (3-2):

إذا كان سحب العينات الطبقية مستقلاً عندئذ:

$$V(\overline{y}_{st})=\sum_{h=1}^{L}W_h^2.V(\overline{y}_h)$$

حيث $V(\overline{y}_h)$ هو تباين $\overline{y}_h$ فوق عيّنات متكررة من الطبقة h.

الإثبات:

بما أن $\overline{y}_{st}=\sum_{h=1}^{L}W_h.\overline{y}_h$

عندئذ نجد أن $\overline{y}_{st}$ هي دالة خطية في المقادير بترجيحات ثابتة W_h. وبالتالي يمكن اقتباس النتيجة الإحصائية المتعلقة بتباين دالة خطية.

$$V(\overline{y}_{st})=\sum_{h=1}^{L}W_h^2.V(\overline{y}_h)+2\sum_{h=1}^{L}\sum_{J>h}^{L}W_hW_j\operatorname{cov}(\overline{y}_h,\overline{y}_j)$$

والحد الثاني من الطرف الأيمن يكون معدوماً لأن سحب العيّنات الطبقية يتمتع بصفة الاستقلالية (فالتغاير معدومة) ومنه:

$$V(\overline{y}_{st})=\sum_{h=1}^{L}W_h^2.V(\overline{y}_h)$$

مبرهنة (3-3): في معاينة عشوائية طبقية ، يكون تباين التقدير:

$$V(\overline{y}_{st})=\frac{1}{N^2}\sum_{h=1}^{L}N_h(N_h-n_h).\frac{S_h^2}{n_h}$$

$$=\sum_{h=1}^{L}W_h^2.\frac{S_h^2}{n_h}(1-f_h)$$

الإثبات: بما أن تقدير $\overline{y}_h$ غير منحاز لـ $\overline{Y}_h$ فيمكن تطبيق المبرهنة (3-2) ومن مبرهنة سابقة في المعاينة البسيطة وعلى طبقة واحدة نجد:

$$V(\bar{y}_h)=\frac{S_h^2}{n_h}.\frac{N_h-n_h}{N_h}$$

وبالتعويض في نتيجة المبرهنة (2-3) نحصل على

$$V(\bar{y}_{st})=\frac{1}{N^2}\sum_{h=1}^{L}N_h^2.V(\bar{y}_h)=\frac{1}{N^2}\sum_{h=1}^{L}N_h(N_h-n_h).\frac{S_h^2}{n_h}$$

$$=\sum_{h=1}^{L}W_h^2.\frac{S_h^2}{n_h}.(1-f_h)$$

وفيما يلي بعض النتائج لحالات خاصة :

نتيجة (1):

إذا كان كسر المعاينة $\frac{n_h}{N_h}$ مهملاً في جميع الطبقات فإن:

$$V(\bar{y}_{st})=\frac{1}{N^2}\sum_{h=1}^{L}\frac{N_h^2.S_h^2}{n_h}=\sum_{h=1}^{L}\frac{W_h^2.S_h^2}{n_h}$$

وهي العلاقة المناسبة عندما يكون إهمال عامل التصحيح ممكناً.

نتيجة (2):

في حالة الحصص المتناسبة ، نعوض:

$n_h=\frac{n.N_h}{N}$ وبالتالي:

$$V(\bar{y}_{st})=\sum_{h=1}^{L}\frac{N_h}{N}.\frac{S_h^2}{n}.\left(\frac{N-n}{N}\right)$$

$$=\frac{(1-f)}{n}\sum_{h=1}^{L}W_h.S_h^2$$

نتيجة (3):

إذا كانت المعاينة تناسبية وكان للتباينات في جميع الطبقات القيمة S_W^2 نفســـــها، فنحصل على النتيجة البسيطة التالية:

$$V(\bar{y}_{st})=\frac{S_W^2}{n}\left(\frac{N-n}{N}\right)$$

مبرهن (3-4):

إذا كان $\hat{Y}_{st} = N\bar{y}_{st}$ هو تقدير لمجموع المجتمع Y ، فعندئذ.

$$V(\bar{y}_{st}) = \sum_{h=1}^{L} N_h (N_h - n_h) . \frac{S_h^2}{n_h}$$

وهذا ينتج مباشرة من المبرهنة (3-2):

مثال:

يبين الجدول التالي تعداوي 1920 ,1930 لسكان 64 من المدن الكبرى في أمريكا بالآلاف ، وقد حصلنا على البيان الإحصائي بأخذ المدن التي يأتي ترتيبها بين 65 ,68 في أمريكا وفقا لعدد سكانها في عام 1920. وقد رتبنا المدن في طبقتين، الأولى تحوي المدن الست عشرة الأكبر وتحوي الثانية المدن الباقية.

n_{hi} حجم عام 1920				y_{hi} حجم عام 1930			
الطبقة				الطبقة			
1	2			1	2		
797	314	172	121	900	364	209	113
773	298	172	120	822	317	183	115
748	296	163	119	781	328	163	123
734	258	162	118	805	302	253	154
588	256	161	118	670	288	232	140
577	243	159	116	638	291	260	119
507	238	153	116	573	253	201	130
507	237	144	113	634	291	147	127
457	235	138	113	578	308	292	100
438	235	138	110	487	272	164	107
415	216	138	110	442	284	143	114
401	208	138	108	451	255	169	111
387	201	136	106	459	270	139	163
381	192	132	104	464	214	170	116
324	180	130	101	400	195	150	123
315	179	126	100	366	260	143	134

(حيث المدن مذكورة بالترتيب نفسه في كل من العامين).

وكانت المجاميع و مجاميع المربعات الموافقة

طبقة	1920		1930	
	$\sum x_{hi}$	$\sum x_{hi}^2$	$\sum y_{hi}$	$\sum y_{hi}^2$
1	8349	3756619	10070	7145450
2	7941	1474871	9498	2141720

وسنقدر تعداد السكان عام 1930 في جميع المدين الـ 64 من عينة حجمها 24.

عين الخطأ المعياري للمجموع المقدر في حالة.

1- عينة عشوائية بسيطة.

2- عينة عشوائية طبقية بحصص متناسبة.

3- عينة عشوائية طبقية بـ 12 وحدة من كل طبقة وهذا المجتمع شبيه بمجتمعات أنواع عديدة من المشروعات التجارية من حيث إن بعض الوحدات (المدن الكبير) تسهم بشكل أكبر بكثير في المجموع الكلي وتظهر قدرا من التغير أكبر بكثير من الوحدات الباقية. وسنستخدم هنا بيانات 1930 فقط، وستظهر بيانات 1920 في مثال لاحق.

وفيما يتعلق بعدد السكان عام 1930 نجد.

$$Y=19568 \quad ; \quad S^2=52448$$

الحل:

1- من أجل المعاينة العشوائية البسيطة:

$$V(\hat{Y})=\frac{N^2 S^2}{n}.\frac{N-n}{N}=\frac{(64)^2\,(52448)}{24}\left(\frac{40}{64}\right)$$

$$=5594453$$

والخطأ المعياري $\sigma(\hat{Y})=2365$

2- في كل من الطبقتين ، نجد أن التباين

$$S_1^2=53843 \quad ; \quad S_2^2=5581$$

ولدينا في الحصص المتناسبة $n_2=18 \quad , \quad n_1=6$

$$V(\hat{Y})=\frac{N-n}{n}\sum_{h=1}^{2}N_h S_h^2$$

$$=\frac{40}{24}\left[(16)(53843)+(48)(5581)\right]$$

$$=1882293$$

والخطأ المعياري $\sigma(\hat{Y})=1372$

3- من أجل $n_1=n_2=12$ نجد:

$$V(\hat{Y})=\sum_{h=1}^{2}N_h\,(N_h-n_h).\frac{S_h^2}{n_h}$$

$$=\frac{(16)(4)(53843)}{12}+\frac{(48)(36)(5581)}{12}=1090827$$

$$\Rightarrow\sigma(\hat{Y})=1044$$

نجد في هذا المثال أن الحجموم المتساوية للعينتين من الطبقتين أكـــثر دقـــة مـــن الحصص المتناسبة وكلاهما أفضل بكثير من المعاينة العشوائية البسيطة.

3-4: تقدير التباين وحدود الثقة:

إذا أخذنا عينة عشوائية بسيطة ضمن كل طبقة ، فإن التقدير غير المنحاز للتبـــاين S_h^2 يكون:

$$s_h^2=\frac{1}{n_h-1}\sum_{i=1}^{n_h}(y_{hi}-\bar{y}_h)^2$$

وبالتالي يكون لدينا المبرهنة التالية:

مبرهنة (3-5):

في المعاينة العشوائية الطبقية يكون التقدير التالي تقديرا غير منحاز لتباين $\bar{y}_{st}$

$$V(\bar{y}_{st})=S^2(\bar{y}_{st})\frac{1}{N^2}\sum_{h=1}^{L}N_h\,(N_h-n_h)\frac{s_h^2}{n_h}$$

ويمكن كتابة هذه العبارة بشكل آخر مناسب للأعمال الحسابية

$$S^2(\bar{y}_{st})=\sum_{h=1}^{L}\frac{W_h^2}{n_h}s_h^2-\sum_{h=1}^{2}\frac{W_h s_h^2}{N}$$

ويمثل الحد الثاني في الطرف الأيمن التخفيض العائد لعامل التصحيح. ولحساب هذا التقدير، يجب أن يكون لدينا وحدتان على الأقل، مسحوبتان من كل طبقة.

والعلاقات الخاصة بحدود الثقة تكون كما يلي:

بما يتعلق بمتوسط المجتمع : $\left[\bar{y}_{st}-tS(\bar{y}_{st}),\bar{y}_{st}+tS(\bar{y}_{st})\right)$

وبما يتعلق بمجموع المجتمع : $\left[N\bar{y}_{st}-tNS(\bar{y}_{st}),N\bar{y}_{st}+tNS(\bar{y}_{st})\right]$

وتفترض هاتان العلاقتان أن $\bar{y}_{st}$ يتوزع طبيعيا. وأن $S(\bar{y}_{st})$ محدد تحديدا جيدا، بحيث يمكن قراءة العامل t من جداول التوزيع الطبيعي.

وإذا قدمت كل طبقة عددا قليلا من درجات الحرية، فإنه الطريقة المعتادة لحساب خطأ العينة الموافق لكمية مثل $S(\bar{y}_{st})$ هي أن نقرأ القيمة t من جداول توزيع ستودنت بدلا من جدول التوزيع الطبيعي وبصورة عامة يكون توزيع $S(\bar{y}_{st})$ من التعقيد بحيث لا يسمح بتطبيق دقيق لهذه الطريقة. والطريقة التقريبية لتخصيص عدد فعال من درجات الحرية لـ $S(\bar{y}_{st})$ هي كما يلي:

$$S(\bar{y}_{st})=\frac{1}{N^2}\sum_{h=1}^{L}g_h S_h^2\ ;\ g_h=\frac{N_h(N_h-n_h)}{n_h}$$

والعدد الفعال من درجات الحرية n_e هو:

$$n_e=\frac{\left(\sum_{h=1}^{L}g_h S_h^2\right)^2}{\sum_{h=1}^{L}\frac{g_h^2.S_h^4}{n_h-1}}$$

وتقع قيمة n_e دائما بين أصغر قيم الكميات (n_h-1) وبين مجموع هذه الكميات. ويأخذ التقريب بالحسبان حقيقية أن s_h^2 يمكن أن يتغير من طبقة إلى طبقة.

ونحتاج هنا إلى الفرض بأن المتغيرات y_{hi} تتوزع وفق التوزيع الطبيعي لأن التقريب ويعتمد على نتيجة أن تباين s_h^2 هو $\frac{2\sigma_h^4}{n_h-1}$. وإذا كان لتوزيع y_{hi} تفرطح

إيجابي فسيكون تباين s_h^2 أكبر من ذلك. وستبالغ علاقة n_e في تقدير العدد الفعال من درجات الحرية.

5-3: المحاصة المثلى:

في المعاينة الطبقية يختار المعاين قيم حجوم العينة n_h في الطبقات المتتالية، وقد يختارها بحيث تجعل $V(\bar{y}_{st})$ أصغر ما يمكن من أجل تكلفة محددة للحصول على العينة، أو بحيث تجعل التكلفة أصغر ما يمكن من أجل قيمة محدده لـ $V(\bar{y}_{st})$. وأبسط دالة تكلفة هي من الشكل

$$C = C_0 + \sum_{h=1}^{L} c_h n_h \quad \text{(التكلفة)}$$

أي تكون التكلفة ضمن كل طبقة متناسبة مع حجم العينة، إلا أن تكلفة وحدة المعاينة C_h يمكن أن تتغير من طبقة لأخرى. ويمثل الحد C_o التكلفة الابتدائية . وتكون دالة التكلفة هذه مناسبة عندما يكون الشيء الرئيسي في التكلفة هي أخذ القياسات في كل وحدة معاينة وإذا كانت التكاليف الانتقال بين الوحدات كبيرة فإن الدراسات الرياضية والتجريبية تقترح أن أفضل تمثيل لتكاليف الانتقال يكون بوسلطة العبارة:

$\sum_{h=1}^{L} t_h \sqrt{n_h}$ حيث t_h يمثل معدل تكلفة الانتقال للوحدة الواحدة وتعد هنا دالة التكلفة الخطية C فقط.

مبرهنة (3-6):

في معاينة عشوائية طبقية مع دالة التكلفة خطية C يكون تباين تقدير المتوسط أصغر ما يمكن من أجل تكلفة محددة C . وتكون التكلفة C أصغر ما يمكن من أجل تباين محدد $V(\bar{y}_{st})$ عندما يتناسب n_h مع $\frac{W_h s_h}{\sqrt{c_h}}$

الإثبات:

لدينا $\quad C = C_0 + \sum_{h=1}^{L} C_h n_h \quad$ ولدينا

$$V = V(\bar{y}_{st}) = \sum_{h=1}^{L} \frac{W_h^2 . S_h^2}{n_h}(1-f_h)$$

$$= \sum_{h=1}^{L} \frac{W_h^2}{n_h} S_h^2 - \sum_{h=1}^{L} \frac{W_h^2 . S_h^2}{N_h}$$

والمشكلة هنا: (1) إما اختيار n_h بحيث يكون V أصغر ما يمكن من أجل C محددة أو (2) اختيار n_h بحيث تكون C أصغر ما يمكن من أجل V محددة.

ويتفق أن يكون للمشكلتين الحل نفسه إذا استثنينا الخطوات الأخيرة، فاختيار المقادير n_h بحيث بجعل V أصغر ما يمكن من أجل C مثبتة، أو جعل C أصغر ما يمكن من أجل V مثبت يكافئ جعل الجداء:

$$V'C' = \left(V + \sum_{h=1}^{L} \frac{W_h^2 . S_h^2}{N_h}\right)(C - C_0)$$

$$= \left(\sum_{h=1}^{L} \frac{W_h^2 S_h^2}{n_h}\right)\left(\sum_{h=1}^{L} C_h n_h\right)$$

أصغر ما يمكن.

وقد لاحظ sturt أنه يمكن بسهولة جعل العلاقة ألآخيرة أصغر ما يمكن وذلك باستخدام متراجحة كوشي – شوارتز:

فإذا كانت a_h , b_h مجموعتين من L من الأعداد الموجبة، فتأتي هذه المتراجحة من المطابقة.

$$\left(\sum_{h=1}^{L} a_h^2\right)\left(\sum_{h=1}^{L} b_h^2\right) - \left(\sum_{h=1}^{L} a_h b_h\right)^2 = \sum_{i}\sum_{j>i}(a_i b_j - a_j b_i)^2$$

حيث ينتيج من هذه العلاقة أن

$$\left(\sum_{h=1}^{L} a_h^2\right)\left(\sum_{h=1}^{L} b_h^2\right) \geq \left(\sum_{h=1}^{L} a_h b_h\right)^2$$

وتتحقق المساواة إذا كانت $\frac{b_h}{a_h}$ ثابتة من أجل جميع قيم h ومنه بأخذ.

$$a_h = \frac{W_h S_h}{\sqrt{n_h}} ; b_h = \sqrt{C_h n_h} ; a_h . b_h = W_h S_h . \sqrt{C_h}$$

فإن المتراجحة أعلاه تعطي :

$$V'C' = \left(V + \sum_{h=1}^{L} \frac{W_h^2 . S_h^2}{N_h}\right)\left(\sum_{h=1}^{L} C_h n_h\right) = \left(\sum_{h=1}^{L} a_h^2\right)\left(\sum_{h=1}^{L} b_h^2\right)$$

$$\geq \left(\sum_{h=1}^{L} W_h S_h . \sqrt{C_h}\right)^2$$

وهكذا فإن لا يوجد اختيار للأعداد n_h يمكن أن يجعل $V'C'$ أصغر من

$$\sum_{h=1}^{L} W_h S_h . \sqrt{C_h}$$

وتقع النهاية الصغرى عندما يكون:

$$\frac{b_h}{a_h} = \frac{n_h \sqrt{C_h}}{W_h S_h} = \text{ثابت}$$

وبدلالة حجم العينة الكلي n_h في طبقة ، نجد:

$$\frac{n_h}{n} = \frac{W_h S_h / \sqrt{C_h}}{\sum_{h=1}^{L} (W_h S_h / \sqrt{C_h})} = \frac{N_h S_h / \sqrt{C_h}}{\sum_{h=1}^{L} (N_h S_h / \sqrt{C_h})}$$

وتقود هذه المبرهنة إلى القواعد الإجرائية التالية: في طبقة معينة، خذ عينة أكبر إذا كانت:

1- الطبقة أكبر

2- التغيرات الداخلية في الطبقة أكبر.

3- المعاينة من الطبقة أرخص.

ونحتاج إلى خطوه إضافية لإتمام المحاصة، إذ تعطي المعادلة الأخيرة قيمة n_h بدلالة n. ولكننا لا نعرف بعد قيمة h، ويتوقف الحل على ما إذا كانت العينة قد اختيرت بحيث تواجه تكلفة إجمالية محددة C أو تعطي تباينا محددا V لـ $\bar{y}_{st}$. وإذا كانت التكلفة مثبته، فنعوض القيم المثلى لـ n_h في دالة التكلفة C ثم نحل لإيجاد n حيث يعطي ذلك:

$$n=\frac{(C-C_0)\sum_{h=1}^{L}(N_h S_h/\sqrt{C_h})}{\sum_{h=1}^{L}(N_h S_h.\sqrt{C_h})}$$

وإذا كان V مثبتا ، فتعوض القيمة المثلى n_h في العلاقة الخاصة: $V(\overline{y}_{st})$ لنجد:

$$n=\frac{\left(\sum_{h=1}^{L}W_h S_h\sqrt{C_h}\right)\sum W_h S_h/\sqrt{C_h}}{V+\frac{1}{N}\sum_{h=1}^{L}W_h S_h^2}$$

حيث $W_h=\frac{N_h}{N}$.

وتبرز حالة خاصة مهمة إذا كان : $C_h=C$ أي إذا بقيت التكلفة لكل وحده نفسها في جميع الطبقات ، فالتكلفة تصبح $C=C_0+Cn$، وتصبح المحاصة المثلى في حالة تكلفة مثبتة هي المحاصة المثلى من أجل حجم عينة ثابت وتكون النتيجة في هذه الحالة الخاصة كما يلي:

مبرهنة (7-3):

في معاينة عشوائية طبقية، يكون $V(\overline{y}_{st})$ أصغر ما يمكن من أجل كلي مثبت للعينة إذا كان

$$n_h=n.\frac{W_h S_h}{\sum_{h=1}^{L}W_h S_h}=n.\frac{N_h S_h}{\sum_{h=1}^{L}N_h S_h}$$

وتدعى هذه المحاصة أحيانا محاصة قيمان.

ونحصل على علاقة التباين الأصغري مع n مثبت، بتعويض قيمة n_h الأخيرة في العلاقة العامة المتعلقة بـ $V(\overline{y}_{st})$ وتكون النتيجة:

$$V\min(\overline{y}_{st})=\frac{\left(\sum_{h=1}^{L}W_h S_h\right)^2}{n}-\frac{\sum_{h=1}^{L}W_h S_h^2}{N}$$

ويمثل الحد الثاني في الطرف الأيمن عامل التصحيح.

6-3: الدقة النسبية لمعاينة عشوائية طبقية ومعاينة عشوائية بسيطة:

إذا استخدمت طريقة التقسيم إلى طبقات بمهارة فإنها ستنتج على الدوام، تقريبا ، تباينا لتقدير المتوسط أو تقدير المجموع، أصغر من التباين الذي تعطيه العينة العشوائية البسيطة المقابلة وعلى أي حال، فإنه ليس صحيحا أن أي عينة عشوائية طبقية تعطي تباينا أصغر مما تعطيه العينة العشوائية البسيطة. وإذا كانت قيم n_h بعيده عن كونها مثلى ، فقد يكون للمعاينة الطبقية تباين أعلى، وفي الحقيقية ، وكما سنبين فإنه في حالة حجم كلي مثبت للعينة، يمكن حتى للتقسيم إلى طبقات مع محاصة مثلى، أن يعطي تباينا أعلى ، علما بأن هذه النتيجة تبدو نوعا من الفضول الأكاديمي أكثر مما هي شيء يحتمل حدوثه في الممارسة العملية. وسوف نقدم في هذه الفقرة مقارنة بين العينة العشوائية البسيطة والعينة العشوائية على إظهار كيفية إنجاز أو كسب العائد لاستخدام طريقة التقسيم إلى طبقات. ونرمز لتباينات تقديرات المتوسط بـ

V_{ra} , V_{ProP} , V_{opt} على الترتيب.

مبرهنة (3-8):

إذا تجاهلنا الحدود $\frac{1}{N_h}$ بالمقارنة مع الواحد يكون

$$V_{opt} \le V_{\text{ProP}} \le V_{ran}$$

وحيث تتم المحاصة المثلى من أجل n مثبته

الإثبات:

لدينا الفقرات السابقة:

$$V_{ran} = (1-f).\frac{S^2}{n}$$

$$V_{\text{ProP}} = \frac{(1-f)}{n}\sum_{h=1}^{L} W_h S_h^2 = \frac{\sum_{h=1}^{L} W_h S_h^2}{n} - \frac{\sum_{h=1}^{L} W_h S_h^2}{N}$$

$$V_{opt} = \frac{\left(\sum_{h=1}^{L} W_h S_h\right)^2}{n} - \frac{\sum_{h=1}^{L} W_h S_h^2}{N}$$

ومن المطابقة الجبرية المعتادة لتحليل تباين المجتمع المقسم إلى طبقات نحصل على:

$$(N-1)S^2=\sum_h\sum_i(y_{hi}-\bar{Y})^2$$
$$=\sum_h\sum_i(y_{hi}-\bar{Y}_h)^2+\sum N_h(\bar{Y}_h-\bar{Y})^2$$
$$=\sum_h(N_h-1)S_h^2+\sum_h N_h(\bar{Y}_h-\bar{Y})^2$$

وبما أن الحدود التي تحوي $\frac{1}{N_h}$ مهملة وبالتالي أيضا الحدود التي تحوي $\longrightarrow \frac{1}{N}$، فالعلاقة الأخيرة تعطي:

$$S^2=\sum_{h=1}^{L}W_hS_h^2+\sum_{h=1}^{L}W_h(\bar{Y}_h-\bar{Y})^2$$

ومنه:

$$V_{ran}=(1-f)\frac{S^2}{n}=\left(\frac{1-f}{n}\right)\sum_{h=1}^{L}W_hS_h^2+\left(\frac{1-f}{n}\right)\sum_{h=1}^{L}W_h(\bar{Y}_h-\bar{Y})^2$$
$$=V_{\text{ProP}}+\frac{(1-f)}{n}\sum_{h=1}^{L}W_h(\bar{Y}_h-\bar{Y})^2$$

ومن تعريف V_{opt} يجب أن يكون $V_{\text{ProP}}>V_{oPt}$ والفرق بينهما:

$$V_{\text{ProP}}-V_{oPt}=\frac{1}{n}\left[\sum_{h=1}^{L}W_hS_h^2-\left(\sum_{h=1}^{L}W_hS_h\right)^2\right]$$
$$=\frac{1}{n}\left[\sum_{h=1}^{L}W_h\left(S_h-\bar{S}\right)^2\right]$$

حيث $\bar{S}=\sum_{h1}^{L}W_hS_h$ يمثل المتوسط المرجح للمقادير S_h ومع إهمال الحدود في $\frac{1}{N_h}$ نجد:

$$V_{ran}=V_{oPt}+\frac{1}{n}\sum_{h=1}^{L}W_h(S_h-\bar{S})^2$$
$$+\frac{(1-f)}{n}\sum_{h-1}^{L}W_h(\bar{Y}_h-\bar{Y})^2$$

وخلاصة القول، إن العلاقة الأخيرة تظهر أن التباين يتناقص وفق مركبتين، عندما ننتقل من المعاينة العشوائية البسيطة إلى المحاصة المثلى، وتأتي المركبة الأولى من حذف الفروق بين متوسطات الطبقات، كما تأتي المركبة الثانية من حذف التأثيرات الناتجة عن الفروق بين الانحرافات المعيارية للطبقات. وتمثل المركبة الثانية الفرق في التباين بين المحاصة المثلى والمحاصة التناسبية. وإذا لم يكن ممكنا إهمال الحدود في $\frac{1}{N_h}$ فإن تعويض S^2 يقود إلى النتيجة التالية:

$$V_{ran}=V_{\text{Pro}P}+\frac{1-f}{n(N-1)}\left[\sum_{h=1}^{L}N_h(\bar{y}_h-\bar{Y})^2-\frac{1}{N}\sum_{h=1}^{L}(N-N_h)S_h^2\right]$$

بدلا من العلاقة السابقة في V_{ran}.

ومنه فإن المعاينة العشوائية التناسبية تعطي تباينا أعلى من المعاينة العشوائية البسيطة، إذا كان:

$$\sum_{h=1}^{L}N_h(\bar{Y}_h-\bar{Y})<\frac{1}{N}\sum_{h=1}^{L}(N-N_h)\,S_h^2$$

ورياضيا، يمكن أن يحدث هذا، فلنفترض أن جميع المقادير S_h^2 تساوي S_W^2 بحيث تكون المحاصة التناسبية مثلى بالمعنى النيماني للكلمة. فعندئذ العلاقة الأخيرة تصبح:

$$\sum_{h=1}^{L}N_h(\bar{Y}_h-\bar{Y})^2<(L-1)\,S_W^2$$

$$\frac{\sum_{h=1}^{L}N_h(\bar{Y}_h-\bar{Y})^2}{L-1}<S_W^2$$

وأولئك الذين يألفون تحليل التباين سيتعرفون على حقيقة أن هذه العلاقة تتضمن كون متوسط مربعات ما بين الطبقات أصغر من متوسط مربعات ما ضمن الطبقات، أي أن النسبة F أقل من الواحد.

7-3: تقدير حجم العينة في حالة معلومات البيانات المتصلة:

رأينا كيف تحدد n تحت محاصة مثلى وسنبحث الآن في تقدير h لأي محاصة، فلقد افترضنا أن هناك للتقدير تباينا محددا V فنحدد بدلا من هامش الخطأ d حيث

$V=\left(\frac{d}{t}\right)^2$ و t تمثل قيمة المتغير الطبيعي الموافقة للاحتمال المسموح به لحادثـــة تجـــاوز الخطأ الفعلي للهامش المرغوب.

ليكن s_h تقدير لـــ S_h و $n_h=\omega_h n$ حيث ω_h مقادير تم اختيارها. وبدلالة هـــذه الحدود يكون:

$$V(\bar{y}_{St})=\frac{1}{n}=\sum_{h=1}^{L}\frac{W_h S_h^2}{w_h}-\frac{1}{N}\sum_{h=1}^{L}W_h S_h^2$$

حيث $W_h=\frac{N_h}{N}$ ومنه:

$$n=\frac{\sum_{h=1}^{L}\frac{W_h^2 S_h^2}{w_h}}{V+\frac{1}{N}\sum_{h=1}^{L}W_h S_h^2}$$

وإذا تجاهلنا عامل التصحيح ، فنجد كتقريب أول

$$n_0=\frac{1}{V}\sum_{h=1}^{L}\frac{W_h^2 . S_h^2}{w_h}$$

وإذا كان $\frac{n_0}{N}$ غير مهمل ، فيمكن حساب n على الشكل

$$n=\frac{n_0}{1+\frac{1}{NV}\sum_{h=1}^{L}W_h S_h^2}$$

وفي حالات خاصة يمكن أن تأخذ هذه العلاقات أشكالاً حسابية أكثر ســــهولة ومنها:

1- حالة محاصة مثلى افتراضية (n مثبت):

$$n=\frac{\left(\sum_{h=1}^{L}W_h S_h\right)^2}{V+\frac{1}{N}\sum_{h=1}^{L}W_h S_h^2}$$

2- حالة محاصة تناسبية:

$$n_0=\frac{\sum_{h=1}^{L}W_hS_h^2}{V};n=\frac{n_0}{1+\frac{n_0}{N}}$$

ومن خلال تقدير مجموع المجتمع، حيث إذا كان V يمثل القيمة المرغو بـــه لـــــ $V(\hat{Y}_{st})$ فإن العلاقات الرئيسية تصبح كالتالي:

فالشكل العام :

$$n=\frac{\sum_{h=1}^{L}\frac{N_h^2S_h^2}{w_h}}{V+\sum_{h=1}^{L}N_hS_h^2}$$

وفي حالة محاصة مثلى افتراضية:

$$n=\frac{\left(\sum_{h=1}^{L}N_hS_h\right)^2}{V+\sum_{h=1}^{L}N_hS_h^2}$$

وفي حالة محاصة متناسبة:

$$n_0=\frac{N}{V}\sum_{h=1}^{L}N_hS_h^2 \quad ; n=\frac{n_0}{1+\frac{n_0}{N}}$$

مثال:

في عينة من كليات جامعات الولايات المتحدة الأمريكية مسحوبة عام 1946 مـــن قبل مكتب التربية وذلك لتقدير عدد المسجلين للعام الدراسي 47-1946 ، والمسألة هنا تتعلق بمجتمع يتضمن 169 من الكليات ودور المعلمين ، وقد رتبت هـــذه في ســـبع طبقات، سنهمل طبقة صغيرة منها، وقد شكلت الطبقات الخمس الأولى وفقا لحجـــم المعهد، وتضمنت السادسة كليات البنات فقط، وحسبت s_h (تقديرات الـــ S_h) من نتائج العام الدراسي 1944-1943. وقد استخدم تقسيم "أمثل" للطبقات مبني علـى هذه القيم لـــ S_h.

وكان الهدف هو معامل اختلاف مقــــداره 0.05 في تقديــر العــدد الإجمــالي للمسجلين. وفي عام 1943 كان عدد المسجلين الإجمالي لهذه المجموعة من الكليــــات 56472. وهكذا يكون الخطأ المعياري المرغوب 2824 = (56472) (0.05) وبالتالي فالتبــاين المرغوب.

$$V=(2824)^2=7974976$$

وقد يكون هناك اعتراض مفادة أن التسجيل في 1946 سيكون أكبر ممـــا هـــو في 1943، وأن هامشا يجب أن يترك لهذه الزيادة، وفي الواقع ، فإن الحسابات تفــــترض فقط أن معامل الاختلاف للكلية الواحدة يبقى كما هو في عامي 1943 أو 1946 وهــو فرض قد لا يكون مجانيا للمنطق. ويبين الجدول التالي قيـــم n_h ، S_h ، $N_h S_h$ الـــتي كانت معروفة قبل تحديد n.

والعلاقة المناسبة لتحديد n هي :

$$n=\frac{\left(\sum_{h=1}^{L} N_h S_h\right)^2}{V+\sum_{h=1}^{L} N_h S_h^2}$$

والتي تنطبق على محاصة مثلى تهدف إلى تقدير مجموع.

ومن غير المحتمل أن يكون عامل التصحيح مهملا في مجتمع كهذا لا يحـــوي إلا 196 وحده. وعلى أي حال وبغية التوضيح سنحسب أول تقريب متجاهلين التصحيح وهو:

$$n_0=\frac{\left(\sum_{h=1}^{L} N_h S_h\right)^2}{V}=\frac{(268411)^2}{7974976}=9034$$

جدول بيانات من أجل تقدير حجم عينة

طبقة	N_h	S_h	$N_h S_h$	n_h
1	13	325	4225	9
2	18	190	3420	7
3	26	189	4914	10

	4	42	82	3444	7	
	5	73	86	6278	13	
	6	24	190	4560	10	
	مجاميع	196		26841	56	

ومن الواضح أنها بحاجة إلى التعديل، فمن أجل القيمة الصحيحة لـــ n نجد:

$$n = \frac{n_0}{1 + \frac{1}{V}\sum_{h=1}^{L} N_h S_h^2} = \frac{9034}{1 + \frac{4640387}{7974976}} = 57.1$$

وقد اختير حجم عينة مساو لـــ 56 وتظهر قيم الـــ n_h للطبقات كل بمفردهـا في العمود الأيمن من الجدول أعلاه.

8-3 : المعاينة الطبقية في حالة النسب:

إذا رغبنا في تقدير نسبة بالواحدات في المجتمع والتي تقع في صف معين C. فإننـــا نصل إلى تقسيم نموذجي للطبقات عندما نستطيع أن نضع في الطبقـــة الأولى جميـــع الوحدات التي تقع في C وفي الثانية كل وحدة لا تقع في C، وعند الفشل في ذلـــك نحاول أن نشكل الطبقات بحيث تختلف نسبة الواحدات من الصف C بالقدر الممكـــن من طبقة إلى أخرى ولتكن $P_h = \frac{A_h}{N_h}$ ، $p_h = \frac{a_h}{n_h}$ نسبة الواحدات من الصـــف C في الطبقة h وفي العينة المأخوذة من هذه الطبقة على الترتيب. فتقدير النسبة في كـــامل المجتمع الموافق لمعاينة عشوائية طبقية هو:

$$P_{St} = \sum_{h=1}^{L} \frac{N_h P_h}{N}$$

مبرهنة (8-3):

في المعاينة العشوائية الطبقية يكون تباين P_{st} معطى بالعلاقة:

$$V(P_{St}) = \frac{1}{N^2} \sum_{h=1}^{L} \frac{N_h^2 (N_h - n_h)}{N_h - 1} \cdot \frac{P_h . Q_h}{n_h}$$

الإثبات:

هناك لدينا حالة خاصة من المبرهنة العامة والمتعلقة بتباين المتوسط المقدر، ومن المبرهنة (3-3) لدينا:

$$V(\bar{y}_{St})=\frac{1}{N^2}\sum_{h=1}^{L}N_h(N_h-n_h).\frac{S_h^2}{n_h}$$

وليكن y_{St} متغيرا قيمته 1 عندما تقع الواحدة في C وصفرا فيما عدا ذلك، وبالتالي حسب الفصل الثاني وما يخص النسب نجد أن:

$$S_h^2=\frac{N_h}{N_h-1}P_h.Q_h$$

عندئذ:

$$V(P_{St})=\frac{1}{N^2}\sum_{h=1}^{L}\frac{N_h^2(N_h-n_h)}{N_h-1}.\frac{P_h.Q_h}{n_h}$$

ملاحظة: في جميع التطبيقات، وحتى لو لم يكن عامل التصحيح مهملا، ستكون الحدود التي تحوي $\frac{1}{N_h}$ مهملة عمليا، وبالتالي يمكن استخدام العلاقة الأبسط إلى حد ما.

$$V(P_{St})=\frac{1}{N^2}\sum_{h=1}^{L}N_h(N_h-n_h).\frac{P_h Q_h}{n_h}$$

$$=\sum_{h=1}^{L}\frac{W_h^2.P_h.Q_h}{n_h}(1-f_h)$$

نتيجة (1):

عندما نستطيع تجاهل عامل التصحيح نكتب

$$V(P_{St})=\sum_{h=1}^{L}W_h^2\frac{P_h Q_h}{n_h}$$

نتيجة (2):

في حالة المحاصة التناسبية:

$$V(P_{St})=\frac{N-n}{N}\cdot\frac{1}{nN}\sum_{h=1}^{L}\frac{N_h^2 P_h Q_h}{N_h-1}$$

$$=(1-f)\cdot\frac{1}{n}\sum_{h=1}^{L}W_h\cdot P_h\cdot Q_h$$

ولتقدير التباين من العينة يمكن تعويض $\frac{P_h q_h}{n_h-1}$ بدلا من الكمية المجهولة $\frac{P_h Q_h}{n_h}$ في أي من العلاقات المذكورة أعلاه.

وينتج أفضل اختيار للمقادير n_h التي تجعل $V(P_{st})$ أصغر ما يمكن من المبرهنة العامة في (3-5).

والتباين الأصغري في حالة قيمة مثبة للحجم الكلي للعينة هو $N_h\cdot\sqrt{P_h Q_h}$. وهكذا فإن:

$$n_h=n\frac{N_h\sqrt{P_h Q_h}}{\sum_{h=1}^{L}N_n\sqrt{P_h Q_h}}$$

التباين الأصغري الموافقة لتكلفة مثبة ، حيث

$$C=C_0+\sum_{h=1}^{L}C_h n_h \quad \text{(التكلفة)}$$

$$n_h=n\frac{N_h\cdot\sqrt{P_h Q_h/C_h}}{\sum_{h=1}^{L}N_h\cdot\sqrt{P_h Q_h/C_h}}$$

ويتم إيجاد قيمة n كما في (3-5).

3-9: المكاسب في الدقة في حالة معاينة طبيعية للنسب:

إذا كانت التكاليف على أساس الوحدة الواحدة هي نفسها في جميع الطبقات، فهناك قاعدتا عمل مفيدتان:

a- يكون الكسب في الدقة من المعاينة الطبقية العشوائية فوق المعاينة العشوائية البسيطة صغيرا أو متواضعا ما لم تتغير النسب P_h تغيرا كبيرا من طبقة إلى طبقة.

b- إذا وقعت جميع النسب P_h بين 0.1 و 0.9 فإن كسب المحاصة المثلى في حالة n مثبت فوق المحاصة التناسبية يكون سحبا بسيطا.

ولتوضيح النتيجة الأولى، يقارن الجدول التالي معاينة عشوائية طبقيـــة (محاصـــة تناسبية) مع معاينة عشوائية بسيطة من أجل ثلاث طبقات بالحجم نفسه $W_h=\frac{1}{3}$.

جدول الدقة النسبية للمعاينة العشوائية الطبقية والبسيطة			
الدقة النسبية %	طبيعية	بسيطة	P_h
	$nV(P_{St})\ \ 1-f=\frac{1}{3}\sum_{h=1}^{L}P_hQ_h$	$nv(p)/1-f=PQ$	
103	2433	2500	0.5, 0.5, 0.6
112	2233	2500	0.3, 0.5, 0.7
132	1900	2500	0.2, 0.5, 0.8
174	1433	2500	0.1, 0.5, 0.9

ونجد أربع حالات: الأولى فيها P_h تساوي 0.6, 0.5, 0.4 في الطبقات الثلاث، أمــا الأخيرة (وهي الأكثر تطرفا) ففيها P_h تساوي 0.1, 0.5, 0.9. ويبين العمودان التاليــان جداء تباينات النسبة المقدرة بـ $\frac{n}{1-f}$ ويعطي العمود الأخير الدقة النسبية للمعاينـــة العشوائية الطبقية إلى المعاينة العشوائية البسيطة. والكسب في الدقة هو كسب كبير في الحالتين الأخيرتين فقط. ولمقارنة المحاصتين ، التناسبية والمثلى في حالة n مثبت، سنجد أنه مع إهمال العامل (1-f) .

$$V_{oPt}=\frac{\left(\sum_{h=1}^{L}W_h\sqrt{P_hQ_h}\right)^2}{n}\ ;\ V_{\text{ProP}}=\frac{\sum_{h=1}^{L}W_hP_hQ_h}{n}$$

وهكذا تصبح الدقة النسبية للمحاصة التناسبية فوق المحاصة المثلى هي:

$$\frac{V_{oPt}}{V_{\text{ProP}}}=\frac{\left(\sum_{h=1}^{L}W_h\sqrt{P_hQ_h}\right)^2}{\sum_{h=1}^{L}W_hP_hQ_h}$$

وإذا وقعت جميع المقادير P_h مبين القيمتين P_0 و $1-P_0$ فإننا نهتم بالقيمة الصغرى التي تأخذها الدقة النسبية. وللتبسيط، ، نأخذ طبقتين بحجمين متساويين $W_1=W_2$ فتبلغ الدقة النسبية الصغرى عندما يكون $P_1=\frac{1}{2}$; $P_2=P_0$ وعندئذ تصبح الدقة النسبية:

$$\frac{V_{oPt}}{V_{ProP}}=\frac{\left(0.5+\sqrt{P_0Q_0}\right)^2}{2\left(0.25+P_0Q_0\right)}$$

ويعطي قيم هذه الدالة معطاة في الجدول التالي: حيث نجد أنه في حالة كون $P_0=0.1$ أو كونها مرتفعة إلى الحد 0.9 فإن الدقة النسبية هي 94%. وفي معظم الحالات نكون ميزنا البساطة والترجيح الذاتي للمحاصة التناسبية أكثر من أن نعوض هذه الخسارة البسيطة في الدقة.

جدول الدقة النسبية للمحاصتين المثلى والتناسبية

P_0	0.4, 0.6	0.3, 0.7	0. 2, 0.8	0.1, 0.9	0.05, 0.95
$RP(\%)$	100.0	99.8	98.8	94.1	86.6

وينبغي التنويه بمحدودية هذا المثال، فهو لا يأخذ بالحسبان الفروق بين تكاليف المعاينة في الطبقات المختلفة. وفي بعض المسوح تكون النسب P_h صغيرة جدا، ولكنها تتراوح مثلا بين 0.001 و 0.05 في الطبقات المختلفة. وقد تتحقق هنا مكاسب مرموقه أكثر من التقسيم الأمثل للطبقات.

10-3: تقدير حجم العينة في حالة النسب:

يمكن استنتاج القوانين المتعلقة بتحديد حجم عينة من القوانين الأكثر شمولا في الفقرة (3-7). وليكن V التباين المرغوب لتقدير النسبة P المتعلقة بالمجتمع ككل. فالقوانين المتعلقة بالنوعين الرئيسين، للمحاصة هي كما يلي: تناسبية:

$$n_0=\frac{\sum_{h=1}^{L}W_h p_h q_h}{V};n=\frac{n_0}{1+\frac{n_0}{V}}$$

المثلى المفترضة:

$$n_0=\frac{\left(\sum_{h=1}^{L}W_h\sqrt{p_h q_h}\right)^2}{V},\ n=\frac{n_0}{1+\frac{1}{NV}\sum_{h=1}^{L}W_h p_h q_h}$$

حيث n_0 هو التقريب الأول، الذي يتجاهل عامل التصحيح ، و n هـــي القيمـــة المصححة أخذين بالحسبان عامل التصحيح. وعند اشتقاق هذه القوانين افترضنـــا أن العامل $\frac{N_h}{N_h-1}$ مساو للواحد. وتنطبق كل نتائج هذه الفقرة على تقدير نسبة ما، وإذا كان من المفضل استخدام النسب المئوية، فالعلاقات نفسها ممكنة التطبيق إذا اعتبرنـــا عن P_h , Q_h, V الخ ، على شكل منسب مئوية.

ولتقدير العدد الكلي من وحدات المجتمع التي تقع في الصنف C أي تقدير P فإننـا نضرب كل التباينات بـــ N^2 .

11-3: تمارين غير محلوله:

تمرين (1)

في مجتمع فيه N = 6 ، L = 2 كانت قيم y_{hi} هي 0,1 في الطبقـة (1) و 4 ,6 ,11 في الطبقة (2) ونريد أخذ عينة حجمها 2. والمطلوب:

a- نبين أن المحاصة المثلى الينماينة ، عند التدوير إلى عدد صحيح ، هـي $n_h = 1$ في الطبقة (1) و $n_h = 3$ في الطبقة (2).

b- احسب $\bar{y}_{st}$ لكل عينة ممكنة يمكن سحبها تحت المحاصة المثلى وتحت المحاصـة التناسـبية. وتحقـق مـن أن التقديريـن غـير منحـازين، وبالتـالي أوجـد : $V_{oPt}(\bar{y}_{st})$, $V_{\text{ProP}}(\bar{y}_{st})$ مباشرة.

C- تحقق من أن $V_{oPt}(\bar{y}_{st})$ يتفق مع العلاقة المعطاة لـ $V(\bar{y}_{st})$ مبرهنـة (3-3) ومن أجل $V_{\text{ProP}}(\bar{y}_{st})$ يتفق مع العلاقة المعطاة لـ $V(\bar{y}_{st})$ المعطاة في النتيجة (2) مـن المبرهنة (3-3).

d- استخدم العلاقة $V_{\min}(\bar{y}_{st})$ في المبرهنة (3-7) لحساب $V_{oPt}(\bar{y}_{st})$ والـذي يتضمن خطأ طفيفا لأنه لا يفسح المجال لحقيقة أن الـ n_h مقربة إلى أقـرب عـدد صحيح. وهل تتفق هذه النتيجة جيدا مع القيمة المصححة.

تمرين (2):

نأخذ عينة من مجموعة الأسر في مدينة لتقدير الكمية الوسطية لممتلكات الأسـرة الواحدة والتي يمكن ردها مباشرة إلى معادل نقدي. وقد قسمت الأسر إلى طبقتـين عالية الإيجار ومنخفضة الإيجار. ويعتقد أن مترلا من طبقة الإيجار العالي يحوي نحـو (9) أمثال ما يحويه منـزل من طبقة الإيجار المنخفض من مثل هذه الممتلكات كما يتوقـع أن يكون S_h متناسبا مع الجذر التربيعي لمتوسط الطبقة، ويوجد 4000 من الأسـر في طبقة الإيجار المرتفع و 20000 في طبقة الإئجار المنخفض. والمطلوب.

a- كيف توزيع عينة من 1000 أسرة بين الطبقتين.

b- كيف ينبغي توزيع العينة إذا كان الهدف هو تقدير الفرق بين ممتكلات الأسرة الواحدة في الطبقتين.

تمرين (3):

تبين المعلومات الإحصائية التالية تقسيم جميع المزارع في منطقة إلى طبقات وفقـــا لحجم المزرعة ، ولمتوسط عدد الفدادين من الذرة في المزرعة ضمن كل طبقة:

حجم المزرعة بالفدان	عدد المزارع N_h	متوسط فدأدين الذرة $\bar{Y}_h$	الانحراف المعياري S_h
0-40	394	5.4	8.3
41-80	461	16.3	13.3
81-120	391	24.3	15.1
121-160	334	34.5	19.8
161-200	169	42.1	24.5
201-240	113	50.1	26.0
241→	148	63.8	35.2
المجموع أو المتوسط	2010	2603	

وفي عينة حجمها 100 مزرعة ، احسب حجموم العينات من كل طبقة بحيث:

a- المحاصة التناسبية.

b- المحاصة المثلى.

وقارن دقة كل من هاتين الطريقتين ، بدقة المعاينة العشوائية البسيطة.

تمرين (4):

يقترح معاين أخذ عينة عشوائية طبقية ، وبتوقع أن تكون التكاليف الميدانية وفـــق الصيغة $\sum_{h=1}^{L} C_h n_h$ وكانت تقديراته المسبقة عن الحجمين المناسبين للطبقتين كما يلي:

الطبقة	W_h	S_h	C_h
1	0.4	10	4 $
2	0.6	20	9 $

والمطلوب :

a- عين قيم $\frac{n_1}{n}$, $\frac{n_2}{n}$ والتي ستجعل التكلفة الميدانية الإجمالية أصغر ما يمكن وذلك في حالة قيمة معطاة $V(\bar{y}_{st})$.

b- عين حجم العينة المطلوبة تحت هذه المحاصة التناسبية، كي تجعل $V(\bar{y}_{st})=1$ مع تجاهل معامل التصحيح.

c- كم ستكون التكلفة الإجمالية للعمل الميداني؟؟.

تمرين (5):

قارن القيمتين اللتين نحصل عليهما بـ $V(p_{st})$ تحت المحاصة التناسبية والمحاصة المثلى في حالة حجم عينة مثبت في كل من المجتمعين التاليين. حيث حجموم الطبقات متساوية، ويمكن إهمال معامل التصحيح. وما هي النتيجة العامة التي يوضحها هذان المجتمعان؟

المجتمع (1)		المجتمع (2)	
طبقة	P_h	طبقة	P_h
1	0.1	1	0.01
2	0.5	2	0.05
3	0.9	3	0.10

الفصل الرابع

المعاينة المنتظمة (النمطية)

4-1: وصف المعاينة:

لنفرض أننا رقمنّا الوحدات الـ N في المجتمع ، يترتيب ما من 1 إلى N فلاختيار عيّنة من n من الوحدات ، نأخذ وحدة من الوحدات الـ K الأولى بصورة عشوائية، ثم نختار بعدها بصورة نمطية مرتبة الوحدة الـ K من كل K من الوحدات التالية. فمثلا إذا كان K= 15 وكانت الوحدة الأولى المسحوبة هي ذات الرقم 13 فإن الوحدات التالية في العينة تكون ذوات الأرقام 28 و 43 و 58 وهكذا. حيث اختيار الوحدة الأولى يحدد كامل العينة وندعو هذا النوع من العينة بالعينة المنتظمة كل K وحدة.

وميزات هذه المعاينة عن المعاينة العشوائية البسيطة هي كما يلي:

1- سحب العينة أسهل

2- توفير كبير في الوقت.

3- يبدو من المرجح أن تكون المعاينة المنتظمة أكثر دقة من المعاينة العشوائية البسيطة، وفي الحقيقة فإنها تقسم المجتمع طبقيا إلى n طبقة مؤلفة من الوحدات الـ K الأولى ، الوحدات الـ K الثانية التي تليها وهكذا.. . وذلك يمكننا أن نتوقع كون العينة منظمة بنحو دقة العينة العشوائية الطبقية الموافقة نفسها، مع وحدة واحدة من كل طبقة. والفرق هو أنه في حالة العينة المنتظمة تقع كل الوحدات في الموضع النسبي نفسه في الطبقة بينما يتحدد الموضع ضمن كل طبقة بصورة منفصلة وعشوائية في المعاينة العشوائية الطبقية، وتنتشر العينة المنتظمة بعدالة أكثر فوق المجتمع، وهذه الحقيقة تجعل المعاينة المنتظمة أحيانا مثال العينات منظمة ممكنة من أجل N= 23 و K= 5

رقم العينة النمطية				
I	II	III	IV	V
1	2	3	4	5
6	7	8	9	10
11	12	13	14	15
16	17	18	19	20
21	22	23		

جدول (4-1)

4-2 : الصلة بالمعاينة العنقودية:

هناك طريقة أخرى للنظر إلى المعاينة المنتظمة فمع N= n K تبين أعمدة الجدول (2-4) العينات المنتظمة الممكنة. ويتضح من هذا الجدول أن المجتمع قد قسم إلى K من وحدات المعاينة الكبيرة، وكل منها تحوي n من الوحدات الأصلية. وعملية اختيار عيّنة منتظمة متموضّعة عشوائياً هو مجر اختيار واحدة من وحدات المعاينة الكبيرة هذه بصورة عشوائية وهكذا فإن المعاينة المنتظمة تؤدي أساساً إلى اختيار وحدة معاينة مركبة لتشكل بمفردها مجمل العينة. والعينة المنتظمة هي عيّنة عشوائية بسيطة تتضمن وحدة عنقودية واحدة من مجتمع يتضمن K من الوحدات العنقودية.

جدول (4-2) إنشاء K من العينات المنتظمة

	رقم العينة					
	1	2	....	i	.....	.... K
	y_1	y_2	.....	y_i	.....	y_K
	y_{K+1}	y_{K+2}	....	y_{K+i}	....	y_{2K}
	....	.....	.....	.....	....	.....
	$y_{(n-1)K+1}$	$y_{(n-1)K+2}$	....	$y_{(n-1)K+i}$	.....	y_{nK}
المتوسطات	$\bar{y}_1$	$\bar{y}_2$		$\bar{y}_i$		$\bar{y}_K$

3-4 تباين تقدير متوسط:

هناك عدة صيغ متعلق بتباين $\bar{y}_{sy}$ متوسط عيّنة منتظمة والصيغ الثلاث المعطاة أدناه تنطبق على أي نوع من المعاينة العنقودية يحوي فيها كل من العناقيد n عنصراً وتتألف العينة من عنقود واحد. وفي هذه الحالات نفترض أن N= n K .

فإذا كان N= n K ، فمن السهل التحقيق من أن $\bar{y}_{sy}$ هو تقدير غير منحاز لـ $\bar{Y}$ من أجل عينة منتظمة عشوائية التوّضع.

وفي التحليل التالي يدل الرمز y_{ij} على العنصر j من العينة المنتظمة i حيث $j=1,2,...,n$, $i=1,2,...,k$ ونرمز لمتوسط العيّنة i بـ $\bar{y}_i$.

مبرهن (4-1) :

تباين متوسط العينة المنتظمة هو:

$$V(\bar{y}_{sy})=\frac{N-1}{N}S^2-\frac{K(n-1)}{N}S^2_{Wsy}$$

حيث $S^2_{Wsy}=\frac{1}{K(n-1)}\sum_{i=1}^{K}\sum_{j=1}^{n}(y_{ij}-\bar{y}_i)^2$ يمثل التباين فيما بين الوحدات التي تقع ضمن العينة المنتظمة نفسها. وتشكل مقام هذا التباين K(n-1) وفقاً للقواعد المعتادة في تحليل التباين حيث تسهم كل من العينات الـ k بـ n-1 درجة من الحرية في مجموع المربعات الموجود في البسيط.

الإثبات:

نعلم أن

$$(N-1)S^2=\sum_{i=1}^{K}\sum_{j=1}^{h}(y_{ij}-\bar{y})^2$$

$$n\sum_{i=1}^{K}(\bar{y}_i-\bar{Y})^2+\sum_{i=1}^{K}\sum_{j=1}^{h}(y_{ij}-\bar{y}_i)^2$$

ولكن تباين $\bar{y}_{sy}$ هو بالتعريف .

$$V(\bar{y}_{sy})=\frac{1}{K}\sum_{i=1}^{K}(\bar{y}_i-\bar{Y})^2$$

ومنه

$$(N-1)S^2=nKV(\bar{y}_{sy})+K(n-1)S^2_{Wsy}$$

وحيث N= nK . وبالتالي:

$$V(\bar{y}_{sy})=\frac{N-1}{N}S^2-\frac{K(n-1)}{N}S^2_{Wsy}$$

نتيجة:

يكون متوسط عينة منتظمة أكثر دقة من متوسط عينة بسيطة إذا كان: $S^2_{Wsy}>S^2$.

الإثبات:

إذا كان $\bar{y}$ متوسط عينة بسيطة حجمها n، فإن :

$$V(\bar{y})=\frac{N-n}{N}.\frac{S^2}{n}$$

ومن المبرهنة السابقة نجد أن $V(\bar{y}_{sy})<V(\bar{y})$ إذا كان:

$$\frac{N-1}{N}S^2-\frac{K(n-1)}{N}S^2_{Wsy}<\frac{N-n}{N}.\frac{S^2}{n}$$

أي إذا كان

$$K(n-1)S^2_{Wsy}>\left[N-1-\frac{N-n}{n}\right]S^2$$

$$\Leftrightarrow\quad K(n-1)S^2_{Wsy}>K(n-1)S^2\Leftrightarrow S^2_{Wsy}>S^2$$

وهذه النتيجة المهمة التي تنطبق على المعاينة العنقودية، بصورة عامة، تفيد بأن المعاينة النمطية أكثر دقة من المعاينة العشوائية البسيطة، إذا كان التباين ضمن العينات المنتظمة دقيقة عندما تكون الوحدات ضمن العينة نفسها غير متجانسة، وغير دقيقة عندما تكون متجانسة.

مبرهنة (2-4):

$$V(\bar{y}_{sy})=\frac{S^2}{n}\left(\frac{N-1}{N}\right)\left[1+(n-1)\rho_W\right]$$

حيث ρ_W يمثل معامل الارتباط بين أزواج من الواحدة الموجودة في العينة المنتظمة نفسها ويعرّف بالشكل:

$$\rho_W=\frac{E(Y_{ij}-Y)(Y_{iu}-Y)}{E(Y_{ij}-\bar{\bar{Y}})^2}$$

حيث البسط هو المتوسط فوق جميع الـ $\frac{Kn(n-1)}{2}$ مـــن الأزواج المتمـــيزّة.
والمقام هو المتوسط فوق جميع القيم الـــ N لـــ y_{ij} وبما أن المقام يمثيل $(N-1)S^2$
فنجد:

$$\rho_W = \frac{2}{(n-1)(N-1)S^2} \sum_{i=1}^{K} \sum_{j<u} (y_{ij} - \overline{Y})(y_{su} - \overline{Y})$$

الإثبات:

لدينا

$$n^2 K V(\overline{y}_{sy}) = n^2 \sum_{i=1}^{K} (\overline{Y}_i - \overline{Y})^2$$

$$= \sum_{i=1}^{K} \left[(\overline{Y}_{i1} - \overline{Y}) + (Y_{i2} - \overline{Y}) + ... + (Y_{in} - \overline{Y}) \right]^2$$

ومجموع الحدود المربعة هو مجموع مربعات الانحرافات عن $\overline{Y}$ أي أنه يســـاوي $(N-1)S^2$ وهذا يعطي:

$$n^2 K V(\overline{y}_{sy}) = (N-1)S^2 + 2 \sum_{i=1}^{K} \sum_{j<u} (y_{ij} - \overline{Y})(y_{iu} - \overline{Y})$$

$$= (N-1)S^2 + (n-1)(N-1)S^2 \rho_W$$

$$V(\overline{y}_{sy}) = \frac{S^2}{n} \left(\frac{N-1}{N} \right) \left[1 + (n-1)\rho_W \right]$$

ويوضح ذلك بأن الارتباط الإيجابي بين وحدات العينة نفســـها يضخّـــم تبـــاين متوسط العينة. وقد يكون حتى الارتباط إيجابي صغير تأثير بين العامل (n-1).

وتعبّر المبرهنتان السابقتان عن $V(\overline{y}_{st})$ بدلالة S^2، وبالتالي تربطانـــه بالتبـــاين الموافق لعينة عشوائية بسيطة وهناك مبرهنة تعبر عن $V(\overline{y}_{st})$ بدلالة التباين الموافـــق لعينة عشوائية طبقية تتألف فيها الطبقات من الوحدات الـــ K الأولى، الوحدات الـــ K التالية وهكذا. وفي الرموز المستخدمة تشير j في y_{ij} إلى الطبقة وسنكتب متوســـط الطبقة على الشكل $\overline{Y}_{.j}$.

مبرهنة (3-4):

$$V(\overline{y}_{sy}) = \frac{S^2_{W\,st}}{n} \left(\frac{N-n}{N} \right) \left[1 + (n-1)\,\rho_{W\,st} \right]$$

حيث
$$S_{Wst}^2=\frac{1}{n(K-1)}\sum_{i=1}^{n}\sum_{i=1}^{K}(y_{ij}-\bar{y}_j)^2$$
وهو التباين بين الوحدات الواقعة في الطبقة نفسها. ونستخدم المقـــام n(K-1) لأن كلاً من الطبقات الـــ n نفسهم بـــ (K-1) درجة من الحرية. وكذلك:
$$\rho_{Wst}=\frac{E(Y_{ij}-Y_{.j})(Y_{iu}-Y_{.u})}{E(Y_{i.j}-\bar{Y}_{.j})^2}$$
وهذه الكمية هي معامل الارتباط بين انحرافات أزواج المفردات موجودة ضمـــن العينة المنتظمة نفسها، عن متوسطات الطبقات
$$\rho_{Wst}=\frac{2}{n(n-1)(K-1)}\sum_{i=1}^{K}\sum_{j<u}\frac{(Y_{ij}-Y_{.j})(Y_{iu}-\bar{y}_{.u})}{S_{Wst}^2}$$
والإثبات

مشابه لما رأيناه في المبرهنة (2-4)

نتيجة: للعينة المنتظمة دقة العينة العشوائية الطبقية نفسها الموافقة. بوحدة واحـــدة ضمن كل طبقة، إذا كان $\rho_{Wst}=0$ وهذا ناتج عن كون $V(\bar{y}_{st})$ في هذا النوع مـــن العينات العشوائية الطبقية مساوٍ لـــ $V(\bar{y}_{st})=\left(\frac{N-n}{N}\right)\frac{S_{Wst}^2}{n}$

مثال:

متعلق البيان الإحصائي في الجدول (3-4) بمجتمع اصطناعي صغير يفُصح عـــن اتجاه صاعد بثبات تقريباً: لدينا $N=40$, $K=10$, $n=4$. ويمثل كل عدد عينة منتظمة والصفوف هي الطبقات ويوضح المثال الحالة التي يكون فيها الارتباط ضمن العينـــة إيجابيا. فعلى سبيل المثال، يقع كل من الأعداد الأربعة 26, 18, 6, 0 في العينـــة الأولى تحت متوسط الطبقة التي ينتمي إليها العدد. ويبقى هذا صحيحا في العينات المنتظمـــة الخمس الأولى ، مع قليل من الاستثناءات وفي العينات الخمس الأخير يبقى الانحـــراف عن متوسط الطبقة إيجابيا في معظم الحالات. وهكذا تكون الحدود الجدائية في p_{Wst} موجبة في معظمها ومن المبرهنة (3-4) نتوقع أن تكون المعاينة المنتظمة أقل دقـــة مـــن المعاينة العشوائية الطبقية مع وحدة واحدة من كل طبقة.

جدول (3-4) بيان إحصائي لـ 10 عينات منتظمة حيث $n = 4$, $N = K, n= 40$

الطبقة	الأرقام المسلسلة للعينات المنتظمة										متوسط الطبقة
	1	2	3	4	5	6	7	8	9	10	
I	0	1	1	2	5	4	7	7	8	6	4.1
II	6	8	9	10	13	12	15	16	16	17	12.2
III	18	19	20	20	24	23	25	28	29	27	23.3
IV	26	30	31	31	33	32	35	37	38	38	33.1
المجاميع	50	58	61	63	75	71	82	88	91	88	72.7

ونحسب التباين $V(\bar{y}_{sy})$ مباشرة من مجاميع العينات المنتظمة على الشكل:

$$V(\bar{y}_{sy})=\frac{1}{K}\sum_{i=1}^{K}(\bar{Y}_i-\bar{Y})^2=\frac{1}{n^2K}\sum_{i=1}^{K}(n\bar{Y}_i-n\bar{Y})^2$$

$$=\frac{1}{160}\left[(50)^2+(58)^2+....+(88)^2-\frac{(727)^2}{10}\right]=11.63$$

وفي معاينة عشوائية بسيطة أو معاينة طبقية نحتاج إلى تحليل تباين المجتمع إلى ما بين الصفوف وما ضمن الصفوف وهذا مبين في الجدول (4,4) ومنه تكون تباينات تقديرات المتوسطات مستخدمين عينات عشوائية بسيطة وعينات عشوائية طبقية كمايلي:

جدول (4-4) تحليل التباين

التغيرات	درجات الحرية	مجموع المربعات	متوسط المربعات
ما بين الصفوف (الطبقات)	$n-1 = 3$	4828.3	
ما ضمن الطبقات	$n(k-1) = 36$	485.5	$13.49 = S^2_{Wst}$
المجموع	$nK-1 = 39$	5313.8	$136.25 = S^2$

$$V_{ran}=\left(\frac{N-n}{N}\right).\frac{S^2}{n}=\frac{9}{10}.\frac{136.25}{4}=30.66$$

$$V_{st}=\left(\frac{N-n}{N}\right).\frac{S^2_{Wst}}{n}=\frac{9}{10}.\frac{13.49}{4}=3.04$$

ويبين الجدول (4-5) المعلومات الإحصائية نفسها، مع عكس ترتيب الملاحظات في الطبقتين الثانية والرابعة وتأثير ذلك هو جعل ρ_{Wst} سالباً، لأنه يجعل معظم الحدود الجدائية بين الانحرافات عن متوسطات الطبقات سالبة وذلك من أجل أزواج من الملاحظات واقعة في العيّنة المنتظمة نفسها. وفي العينة المنتظمة الأولى مثلاً ، تصبح الانحرافات عن متوسطات الطبقات الآن كما يلي : 4.1- , 4.8 , 5.3 - ,4.9 ومن بين الجداءات الستة لأزواج الانحرافات نجد أن أربعة منها سالبة، وعلى وجه التقريب تنطبق الحالة نفسها على عيّنة منتظمة.

وهذا التغير لا يؤثر في V_{ran} , V_{st} وتؤدي في حالة المعاينة المنتظمة إلى زيادة مثيرة في الدقة.

جدول (4-5) البيان إحصائي في الجدول (4-3) مع عكس الترتيب في الطبقتين IV, II

الطبقة	الأرقام المسلسلة للعينة المنتظمة										متوسط الطبقة
	1	2	3	4	5	6	7	8	9	10	
I	0	1	1	2	5	4	7	7	8	6	4.1
II	17	16	16	15	12	13	10	9	8	6	12.2
III	18	19	20	20	24	23	25	28	29	27	23.3
IV	38	38	37	35	32	33	31	31	30	26	33.1
المجموع	73	74	74	72	73	73	73	75	75	65	72.7

وكما سنرى عند مقارنة مجاميع العينات المنتظمة في كل من الجدولين (4-5) و (4-3) ولدينا الآن:

$$V_{sy}=\frac{1}{160}\left[(73)^2+(74)^2+\ldots+(65)^2-\frac{(727)^2}{10}\right]=0.46$$

4-4: دراسة مجتمعات ذات ترتيب عشوائي:

تستخدم المعاينة المنتظمة أحيانا، لسهولتها، في مجتمعات يكون ترقيم الوحـــدات فيها عشوائيا فعلا. ويكون الأمر كذلك عند معاينة مجموعة بطاقات مرتبة أبجديا وفقا لأسماء الكنية، إذا لم يكن للمفردة التي نقيسها أي علاقة بكنية الشخص. وســوف لا يوجد عندئذ أي اتجاه أو تقسيم إلى طبقات في y_i ونحن نمضي علـــى طـــول هـــذه البطاقات، كما لا يوجد أي ارتباط بين القيم المتجاورة. وفي هذه الحالة، يمكـــن أن نتوقع نوعا من التكافؤ بين المعاينة المنتظمة والمعاينة العشوائية البسيطة، وأن يكون لهـــا التباين نفسه. وليس هذا صحيحا بالضبط في أي مجتمع منته بمفرده مع قيم معطاة لـــ n و K ، على فرض أن V_{sy} المبني على K من درجات الحرية فقط، سيكون غريـــب الأطوار عندما يكون K صغيرا، ويمكن أن يتمخض عن قيمة أكبر أو أصغر مـــن V_{ran} وهناك نتيجتان تبينان أن التباينين هما ، في المتوسط متساويان.

مبرهنة (4-4): لنفرض كل الـــ !N من المجتمعات المشكلة من التباديل الـــ !N لأي مجموعة من الأعداد $y_1, y_2, \ldots, y_N$ وبأخذ المتوسط فوق هذه المجتمعات المنتهية نجد:

$$E(V_{sy}) = V_{ran}$$

ونلاحظ أن V_{ran} يبقى نفسه في جميع التباديل.

وهذه النتيجة تبين أنه إذا أمكن اعتبار ترتيب المفردات في مجتمع معين منته وكأنه مسحوب عشوائيا من التباديل الـــ !N، فعندئذ تكون المعاينة المنتظمـــة مكافئـــة في المتوسط، للمعاينة العشوائية البسيطة.

والطريقة الثانية هي أن نعد المجتمع المنتهي وكأنه مسحوب عشوائيا من مجتمـــع لا نهائي نوني له خواص معينة.

مبرهنة (4-5):

إذا كانت المتغيرات y_i حيث C= 1, 2,..., N مسحوبة عشوائيا من مجتمع نـــوني فيه.

$$E(Y_i - \mu)^2 = \sigma_i^2 \quad ; \; E(Y_i - \mu)(Y_j - \mu) = 0 \;, \; i \neq j$$
$$E(y_i) = \mu$$

عندئذ: $E(V_{sy}) = E(V_{ran})$

والشروط الحاسمة هي أن يكون لجميع المقادير y_i المتوسط μ نفسه أي عـــدم وجود أي اتجاه، وألا يوجد ارتباط خطي بين القيمتين y_i , y_j في نقطتين مختلفيـــتن. ويمكن أن يتغير التباين σ_i^2 من نقطة إلى أخرى في السلسلة.

الإثبات:

في أي مجتمع معين منته لدينا:

$$V_{ran} = \frac{N-n}{Nn} \cdot \frac{\sum_{i=1}^{N} (y_i - \overline{Y})^2}{N-1}$$

والآن:

$$\sum_{i=1}^{N} (y_i - \overline{Y})^2 = \sum_{i=1}^{N} \left[(y_i - \mu) - (\overline{Y} - \mu) \right]^2$$

$$= \sum_{i=1}^{N} (y_i - \mu)^2 - N(\overline{Y} - \mu)^2$$

وبما أن y_i , y_j $(i \neq j)$ غير مرتبطين فلدينا:

$$E(\overline{Y} - \mu)^2 = \frac{1}{N^2} \sum_{i=1}^{N} \sigma_i^2$$

ومنه:

$$EV_{ran} = \frac{N-n}{Nn(N-1)} \left[\sum_{i=1}^{N} \sigma_i^2 - N \frac{\sum_{i=1}^{N} \sigma_i^2}{N^2} \right]$$

$$\Rightarrow EV_{ran} = \frac{N-n}{N^2 n} \cdot \sum_{i=1}^{N} \sigma_i^2$$

من أجل V_{sy}، ولنرمز بـــ $\overline{y}_u$ لمتوسط العينة المنتظمة الـــ u ففي أي مجتمع منتـــه معين لدينا:

$$V_{sy} = \frac{1}{K} \sum_{u=1}^{K} (\overline{y}_u - \overline{Y})^2 = \frac{1}{K} \left[\sum_{u=1}^{K} (\overline{y}_u - \mu)^2 - K(\overline{Y} - \mu)^2 \right]$$

وبالاستناد إلى مبرهنة تباين متوسط عينة غير مرتبطة من مجتمع لانهائي نجد:

$$E(V_{sy}) = \frac{1}{K}\left(\frac{\sum_{i=1}^{N}\sigma_i^2}{n^2} - \frac{K\sum_{i=1}^{N}\sigma_i^2}{N^2}\right) = \frac{N-n}{N^2 n}\sum_{i=1}^{N}\sigma_i^2$$

$$= E(V_{ran})$$

5-4: تمارين:

1- البيان الإحصائي التالي يمثل عدد الشتول في كل قدم من مسكن طولهـــا 2000 قدم. احسب تباين متوسط عينة منتظمة مؤلفة من القدم العشرين من كل عشـــــرين قدماً وقارنه مع التباين في حالة:

a- عينة عشوائية بسيطة.

b- عينة عشوائية طبقية بوحدتين من كل طبقة.

c- عينة عشوائية طبقية بوحدة واحدة من كل طبقة.

في جميع العينات نأخذ $n=10$, $\sum_i (y_i - \bar{Y})^2 = 23.601$

عدد الشتول

الطبقة	1-20 1	21-40 2	41-60 3	61-80 4	81-100 5	101-120 6	121-140 7	141-160 8	161-180 9	161-200 10	مجاميع العينة المنتظمة
				قدم							
1	8	20	26	34	31	24	18	16	36	10	223
2	6	19	26	21	23	19	13	12	8	35	182
3	6	25	10	27	41	28	7	8	29	7	188
4	23	11	41	25	18	18	9	10	33	9	197
5	25	31	30	32	15	29	11	12	14	12	211
6	16	26	55	43	21	24	20	20	13	7	245
7	22	29	34	33	8	33	16	17	18	6	222
8	21	19	56	45	22	37	9	12	20	14	255
9	22	17	39	23	11	32	14	7	13	12	190
10	18	28	41	27	3	26	15	17	24	15	214
11	26	16	27	37	4	36	20	21	29	18	234
12	28	9	20	14	5	20	21	26	18	4	165
13	11	22	25	14	11	43	15	16	16	4	177
14	16	26	39	24	9	27	14	18	20	9	202

15	7	17	24	18	25	20	13	11	6	8	149
16	22	39	25	17	16	21	9	19	15	8	191
17	44	21	18	14	13	18	25	27	4	9	193
18	26	14	44	38	22	19	17	29	8	10	227
19	31	40	55	36	18	24	7	31	8	5	255
20	26	30	39	29	9	30	30	29	10	3	235
المجموع	410	459	674	551	325	528	303	358	342	205	4155

2- رتب مجتمع من 360 أسرة (مرقمة من 1 إلى 360) ترتيبا أبجديا وفقا لكنية معيل الأسرة ووضع في ملف ووقعت التي لم يكن معيلوها من البيض عند الأرقام التالية:

58, 56, 55, 47, 45, 44, 36- 41, 31- 33, 28,, 99-101, 89- 94, 26, 85, 83,
82, 69, 68, 224, 223, 178, 156, 154, 114, 107- 110, 325- 331, 306- 323,
302- 304, 298- 300, 296, 342, 341, 335- 339, 333

قارن دقة 1 إلى 8 عينة منتظمة مع عينة عشوائية بسيطة من الحجم نفسه وذلـك لتقدير نسبة الأسر التي لا يكون معيلها أبيض.

الفصل الخامس

المعاينة العشوائية العنقودية

I– عناقيد متساوية الحجم

II- عناقيد ذات حجوم غير متساوية:

I– عناقيد متساوية الحجم

1-I-5: أسباب المعاينة العنقودية:

في كثير من الدراسات الإحصائية تتألف وحدة المعاينة فيها من فئة أو عنقود من الوحدات الأصغر والتي ندعوها عناصر أو وحدات جزئية.

وهناك سببان رئيسان للمعاينة العنقودية وهي عدم توافر قائمة موثوقة بعناصر المجتمع، أو قد يكون وضع قائمة كهذه مكلفا إلى درجة تجعلها بعيدة المنال . حيث في العديد من الدول لا توجد قوائم كاملة وحديثة للسكان أو للمنازل أو للمزارع.

وباستخدام خرائط للمنطقة يمكننا، على أي حال، تقسيمها إلى وحدات متساوية مثل جادات في المدن أو قطاعات من الأرض حدودها قابلة للتعريف بسهولة في الأجزاء الريفية. وغالبا ما يجري اختيار هذه العناقيد في العديد من الدول المتقدمة لأنها تحل مشكلة وضع قائمة بوحدات المعاينة.

2-I-5: القاعدة البسيطة:

عندما يكون هدف الدراسة هو مقارنة عدد قليل من حجوم محددة للوحدات أو من أنواع محدده منها، فإن المبرهنة التالية توضح لنا العمل.

مبرهنة (1):

إذا كان M_u يمثل الحجم النسبي للوحدة و S_u^2 يمثل التباين بين مجاميع الوحدات و C_u يمثل التكلفة النسبية لقياس وحدة واحدة. عندئذ تكون التكلفة النسبية من أجل تباين محدد، أو التباين النسبي من أجل تكلفة محددة متناسبا مع $\frac{C_u S_u^2}{M_u^2}$، وينطبق هذا على معاينة عشوائية بسيطة يمكن فيها إهمال معامل التصحيح.

الإثبات:

بفرض أن V هو التباين المحدّد لتقدير مجموع المجتمع ، ففي النوع u من الوحدات يكون التقدير $N_u\,\bar{y}_u$ وتباينه $V=\frac{N_u^2\,S_u^2}{n_u}$ حيث $n_u=\frac{N_u^2.\,S_u^2}{V}$ وتكلفة أخذ n_u من الوحدات هي $C_u n_u=\frac{C_u\,N_u^2\,S_{u2}}{V}$ وبما أن $N_u\,M_u$ يبقى ثابتاً في أنواع مختلفة من الوحدات فتكون التكلفة متناسبة مع $\frac{C_u\,S_u^2}{M_u^2}$ ، وعلى الوجه الآخر إذا كانت التكلفة C محددّه فإن $n_n=\frac{C}{C_u}$ ومنه نجد أن V متناسب مع $\frac{C_u\,S_u^2}{M_u^2}$.

نتيجة (1):

إذا عرفنا الدقة النسبية الصافية لوحده بأنها تتناسب عكسيا مع التباين الذين نحصل عليه من أجل الدقة النسبية الصافية متناسبة مع $\frac{M_u^2}{C_u S_u^2}$

نتيجة (2):

في تحليل التباين، غالبا ما تحسب التباينات لوحدات مختلفة الحجوم، على أساس يدعى الأساس المشترك وهو في العادة ذلك الأساس القابل للتطبيق على الوحدة الأصغر. ولوضع التباينات على أساس مشترك نقسم التباين S_u^2 بين مجاميع وحدات حجمها M_u على M_u.

فليكن التباين بين مجاميع الوحدات (على أساس مشترك) يساوي $S_u'^2=\frac{S_u^2}{M_u}$.

فالتكلفة النسبية لأخذ حجم معطى للعينة يساوي $C_u'\sim\frac{C_u}{M_u}$ (حيث ~ يتناسب مع). وبالتالي يمكن عرض المبرهنة (1) كما يلي:

التكلفة النسبية من أجل الدقة نفسها تتناسب مع $\frac{C_u\,S_u^2}{M_u^2}$ تتناسب مع $C_u'.S_u'^2$

الإثبات:

بفرض أن V هو التباين المحدّد لتقدير مجموع المجتمع ، ففي النوع u من الوحدات يكون التقدير $N_u \bar{y}_u$ وتباينه $V = \frac{N_u^2 S_u^2}{n_u}$ حيث $n_u = \frac{N_u^2 . S_u^2}{V}$ وتكلفة أخذ n_u من الوحدات هي $C_u n_u = \frac{C_u N_u^2 S_{u2}}{V}$ وبما أن $N_u M_u$ يبقى ثابتاً في أنواع مختلفة من الوحدات فتكون التكلفة متناسبة مع $\frac{C_u S_u^2}{M_u^2}$ ، وعلى الوجه الآخر إذا كانت التكلفة C محددّه فإن $n_n = \frac{C}{C_u}$ ومنه نجد أن V متناسب مع $\frac{C_u S_u^2}{M_u^2}$.

نتيجة (1):

إذا عرّفنا الدقة النسبية الصافية لوحده بأنها تتناسب عكسياً مع التباين الذين نحصل عليه من أجل الدقة النسبية الصافية متناسبة مع $\frac{M_u^2}{C_u S_u^2}$

نتيجة (2):

في تحليل التباين، غالباً ما تحسب التباينات لوحدات مختلفة الحجوم، على أساس يدعى الأساس المشترك وهو في العادة ذلك الأساس القابل للتطبيق على الوحدة الأصغر. ولوضع التباينات على أساس مشترك نقسم التباين S_u^2 بين مجاميع وحدات حجمها M_u على M_u.

فليكن التباين بين مجاميع الوحدات (على أساس مشترك) يساوي $S_u'^2 = \frac{S_u^2}{M_u}$.

فالتكلفة النسبية لأخذ حجم معطى للعينة يساوي $C_u' \sim \frac{C_u}{M_u}$ (حيث ~ يتناسب مع). وبالتالي يمكن عرض المبرهنة (1) كما يلي:

التكلفة النسبية من أجل الدقة نفسها تتناسب مع $\frac{C_u S_u^2}{M_u^2}$ تتناسب مع $C_u' . S_u'^2$

والدقة النسبية الصافية تتناسب مع $\frac{1}{C'_u \ S'^2_u}$.

وإذا تجاهلنا الفروق في تكاليف أخذ عينة (أي إذا كان C_u ثابتاً، تكون الدقـــة النسبية الصافية للوحدة من تناسبه مع $\frac{1}{S'^2_u}$. ولذلك تكون عوامل أثر التصميـــم في الوحدات المختلفة متناسبة مع $S'^2_u = \frac{S^2_u}{M_u}$.

مثال (5-1):

يقدم بيان جوشون والمتعلق بمسكبة من أغراس الصنوبر الأبيض مثالاً بســـيطاً. فقد احتوت المسكبة ستة صفوف طول كل منها 434 قدماً. وهناك العديد من الطرائق التي يمكن تقسيم المكسبة بموجبها إلى وحدات معاينة. ويبّين الجدول التالي بياناً مـــن أجل أربعة أنواع. من الوحدات. وبما أن المسكبة أحصيت بالكامل فإن البيان يعطـــي قيماً صحيحة للمجتمع. وكانت الوحدات: قدماً واحدة من صف بمفرده، قدمين من صف بمفرده، قدماً واحدة من عرض المسكبة، قدمين من عرض المسكبة.

بيان إحصائي لأربعة أنواع من وحدات المعاينة

بيان إحصائي تمهيدي	نوع الوحدة			
	قدم واحده صف	2 قدم صف	قدم واحد مسكبة	2 قدم مسكبة
M_u الحجم النسبي للوحدة	1	2	6	12
N_u عدد الوحدات في المجتمع	2604	1302	434	217
S^2_u تباين المجتمع لكل وحده	2.537	6.746	23.094	68.558
عدد الأقدام المتتاليـــة الـــتي يمكـــن إحصاؤها في 15 دقيقة	44	62	78	108

وقد افترض في الوحدتين الأولى والثانية أن المعاينة يمكن أن تكون طبقيـــة وفقـــاً للصوف، بحيث يمثل الـــ S_u^2 تباينات ضمن الصفوف، كمـــــا افـــترضت المعاينـــة العشوائية البسيطة في الوحدتين الأخيرتين.

وبما أن التكلفة الرئيسة تقع في تعيين مواضع الوحدات ثم تعدادها فقد قـــــدّرت التكاليف وفقاً لدراسة زمنية (الصف الأخير من الجدول السابق) وفي حالة العينـــــات الأكبر، يمكن تعداد قدر كبير من العينة في 15 دقيقة، لأن الوقت اللازم للتحرك مـــــن وحدة إلى أخرى يصبح أقل.. والكمية المراد تقديرها هي مجموع الأغراس الكلـــي في المسكبة. ووفقاً للمبرهنة (1) يعطي الجدول السابق قيم M_u , S_u^2 والقيم النسبية لـــ C_u معبراً عنها بدلالة الزمن اللازم لتعداد وحدة واحدة هي كما يلي:

	قدم واحد صف	2 قدم صف	قدم واحد مسكبة	2 قدم مسكبة
C_u (في فترة 15 دقيقة)	$\frac{1}{44}$	$\frac{2}{62}$	$\frac{6}{78}$	$\frac{12}{108}$

وبالاستناد إلى نتيجة (1) من المبرهنة (1) تم حساب قيم الدقة النسبية الصافية.

	قدم واحد صف	2 قدم صف	مقدم واحد مسكبة	2 قدم مسكبة
$\frac{M_u^2}{C_u S_u^2}$	$\frac{44}{2.537}=17.3$ 100	$\frac{(4)(62)}{(2)(6.746)}$ =18.38 106	$\frac{(36)(78)}{(3)(23.094)}$ =20.27 117	$\frac{(144)(108)}{(12)(68.558)}$ =18.90 109

والتباينات بين الوحدات، معبراً عنها بدلالة أساس مشترك جديرة بأن ينظر إليــها

أيضاً ، فالقيم $S_u'^2 = \frac{S_u^2}{M_u}$ مطبقة على قدم واحدة من صف ، هي على الترتيب:

5.713 , 3.849 , 3.373 2.537

ونلاحظ أن هذه التباينات تتزايدة فاطراد مع تزايد حجم الوحدة وبما أن الدقـــة

النسبية الصافية متناسبة مع $\frac{1}{C_u S_u^{'2}}$ ، فإن تكلفة أخذ حجم يعطى للعينـــة يجـــب أن

يتناقص في الوحدات الكبيرة، إذا كان لهذه الوحدات أن تثبت اقتصاديتها.

3-I-5 : التباين بدلالة الارتباط ضمن العنقود:

يعبّر أحياناً عن علاقات التباين بدلالة معامل الارتباط ρ بين العناصر الواقعـــة في

العنقود نفسه ، لتكن y_{ij} القيمة الملحوظة للعناصر j ضمن الوحـــدة i وليكـــن Y_i

فنحتاج في المعاينة العنقودية إلى التمييّز بين نوعين من المتوسطات:

المتوسط لكل وحدة . $\bar{Y} = \frac{\sum Y_i}{N}$

والمتوسط لكل عنصر $\bar{\bar{Y}} = \frac{\sum Y_i}{NM} = \frac{\bar{Y}}{M}$

والتباين بين العناصر هو: $S^2 = \frac{\sum_{i,j}(y_{ij} - \bar{\bar{Y}})^2}{NM-1}$

ولقد عرفنا في (3-4) معامل الارتباط العنقودي الداخلي ρ بالشكل:

$$\rho = \frac{E(Y_{ij} - \bar{\bar{Y}})(Y_{iK} - \bar{\bar{Y}})}{E(Y_{ij} - \bar{\bar{Y}})^2} = \frac{2\sum_i \sum_{j<k}(Y_{ij} - \bar{\bar{Y}})(y_{iK} - \bar{\bar{Y}})}{(M-1)(NM-1)S^2}$$

وعدد الحدود E في البسيط هـــو $\frac{NM(M-1)}{2}$ أمـــا E في المقـــام فيســـاوي

$$\frac{(NM-1)S^2}{NM}$$

مبرهنة (2-5):

إذا سُحبت عيّنة عشوائية بسيطة من n عنقوداً كل منها يحوي M عنصراً من الـــ N عنقوداً في المجتمع فعندئذ يكون متوسط العينة لكل عنصر $\bar{\bar{y}}$ تقديراً غير منحاز لـــ $\bar{\bar{Y}}$ بتباين يساوي.

$$V(\bar{\bar{y}})=\frac{1-f}{n}\cdot\frac{NM-1}{M^2(N-1)}S^2\left[1+(M-1)\rho\right]$$

حيث ρ معامل الارتباط ضمن العنقود.

الإثبات:

لنرمز بـــ y_i لمجموع العنقود i و $\bar{y}=\frac{\sum_{i=1}^{n}y_i}{n}$ ورأينا أن $\bar{y}$ هو تقدير غير منحـــاز لـــ $\bar{Y}$ بتباين $V(\bar{y})=\frac{1-f}{n}\cdot\frac{\sum(y_i-\bar{Y})}{N-1}$

إلا أن $\bar{Y}=M\bar{\bar{Y}}$ ، $\bar{y}=M\bar{\bar{y}}$ وبالتالي يكون $\bar{\bar{y}}$ بتباين يساوي

$$V(\bar{\bar{y}})=\frac{1-f}{nM^2}\cdot\frac{\sum(y_i-\bar{Y})^2}{N-1}$$

ولكن:

$$(y_i-\bar{Y})=(y_{i1}-\bar{\bar{Y}})+(y_{i2}-\bar{\bar{Y}})+.......+(y_{iM}-\bar{\bar{Y}})$$

فإذا أخذنا المربعات وجمعنا فوق جميع العناقيد الـــN نجد :

$$\sum_{i=1}^{N}(y_i-\bar{Y})^2=\sum_{i=1}^{N}\sum_{j=1}^{M}(y_{ij}-\bar{\bar{Y}})^2+2\sum_{i=1}^{N}\sum_{P<K}^{M}(y_{ij}-\bar{\bar{Y}})(y_{iK}-\bar{\bar{Y}})$$

$$=(NM-1)S^2+(M-1)(NM-1)\rho S^2$$

$$=(NM-1)S^2\left[1+(M-1)\rho\right] \qquad (*)$$

مستخدمين تعريف ρ . وبالتعويض نجد:

$$V(\bar{\bar{y}})=\frac{1-f}{n}\cdot\frac{NM-1}{M^2(N-1)}S^2\left[1+(M-1)\rho\right]$$

وهذه النتيجة تدل على أنه إذا كان $\rho>0$ يكون العنقود أقل دقة من أجل حجم معطى للعينة. وإذا كان $\rho<0$ كما يحدث أحياناً، فإن العنقود يكون أكثر دقة.

نتيجة:

يمكن إعطاء عبارة بديلة لـ ρ حيث نرمز بـ S_b^2 للتباين بين مجاميع العناقيد على أساس وحدة بمفردها، فعندئذ $\sum_{i=1}^{N}(y_i-\bar{Y})^2=(N-1)M\,S_b^2$ ومنه بالمقارنة مع علاقة هذا المجموع بـ S^2 أعلاه نجد:

$$(N-1)M\,S_b^2=(NM-1)S^2\left[1+(N-1)\rho\right]$$

ومنه :

$$\rho=\frac{(N-1)M\,S_b^2-(NM-1)\,S^2}{(NM-1)(M-1)\,S^2}$$

وعندما تكون الحدود في $\frac{1}{N}$ مهملة يكون:

$$\rho=\frac{S_b^2-S^2}{(M-1)\,S^2}$$

والجدير بالملاحظة هو قيمة متوسط مربعات متضمن العناقيد

$$S_W^2=\sum_{i=1}^{N}\sum_{j=1}^{M}(y_{ij}-\bar{Y}_i)^2/N(M-1)$$

وفي تحليل التباين بتصنيف أحادي لدينا العلاقة

$$(NM-1)S^2=\sum_{i=1}^{N}(y_i-\bar{Y})^2/M+\sum_{i=1}^{N}\sum_{j=1}^{M}(y_{ij}-\bar{y}_i)^2$$

$$=\frac{(NM-1)}{M}S^2\left[1+(M-1)\rho\right]+N(M-1)\,S_W^2$$

ومنه نجد من (*) :

$$S_W^2=\frac{NM-1}{NM}S^2(1-\rho)\approx S^2(1-\rho)$$

ويعد ρ بالتالي مقياس تجانس للعنقود.

4-I-5: المعاينة العنقودية في حالة النسب:

نفترض أنه يمكن تصنيف العناصر الـــ M في أي عنقود إلى صفين، وأن $P_i = \frac{a_i}{M}$ تمثل نسبة العناصر من C في العنقود i ، تؤخذ عينة عشوائية بسيطة من n عنقـــودا ، ونستخدم المتوسط p للنسب الملحوظة p_i في العينة كتقدير لنسبة المجتمع P. وكمـــا ذكرنا في دراسة النسب، نستخدم العلاقة الخاصة بالمتغيرات المستمرة على النسب p_i حيث يعطي:

$$V(P) = \frac{N-n}{Nn} \frac{\sum_{i=1}^{N}(p_i - P)^2}{N-1} = \frac{N-n}{N^2 n}\sum_{i=1}^{N}(p_i - P)^2$$

ومن جهة أجرى، إذا أخذنا عينة عشوائية بسيطة تحـــوي nM مـــن العنـــاصر ، محتسب مبرهنة ذي الحدين نحصل على تباين p وفق الصيغة:

$$V_{brn}(p) = \frac{NM - nM}{NM - 1} \cdot \frac{PQ}{nM} = \frac{N-n}{N} \cdot \frac{PQ}{nM}$$

وكان إذا كان N كبيرا .

وبالتالي يبين عامل أثر التصميم وهو:

$$\frac{V(p)}{V_{bin}(p)} = \frac{M\sum_{i=1}^{N}(p_i - P)^2}{NPQ} \qquad \text{(من أجل N كبيرا)}$$

المتغير النسبي في التباين الذي يعود إلى استخدام العناقيد. والقيم العدديـــة لهـــذا العمل مفيدة في صنع تقديرات تمهيدية لحجم العينة في حالة معاينة عنقودية. ويقـــدر أولا الحجم المطلوب للعينة بالاستناد إلى علاقة التوزيع الحداني، ثم يضرب بالعـــامل للحصول على الحجم الذي سيكون ضروريا في حالة معاينة عنقودية.

وإذا كانت حجوم العناقيد M_i متغيرة فإن التقدير $p=\frac{\sum a_i}{\sum M_i}$ هـــو التقديـــر النسبة وتباينة معطى تقريبا:

$$V(p)=\frac{N-n}{Nn\bar{M}^2}\cdot\frac{\sum_{i=1}^{N} M_i^2\,(p_i-P)^2}{N-1}$$

حيث $\bar{M}=\frac{\sum M_i}{N}$ هو الحجم المتوسط للعنقود.

وإذا قورنت هذه العينة بعينة عشوائية بسيطة من $n\bar{M}$ من العناصر عندئذ نجـــد كتعميم:

$$\frac{V(p)}{V_{bin}(p)}=\frac{\sum_{i=1}^{N} M_i^2(p_i-P)^2}{N\bar{M}PQ}$$

II- عناقيد ذات حجوم غير متساوية:

5-II-1 : وحدات عنقودية ذات حجوم غير متساوية:

في معظم التطبيقات تتضمن الوحدات العنقودية (مثلا: مناطق، مدن، جـــادات مدنية) أعدادا مختلفة من العناصر أو الوحدات الجزئية (وحــدات مســاحة، أســر، أشخاص). وسنعالج ذلك في هذه الفقرة.

ليكن M_i عدد العناصر في الوحده i نعلم من الدراســات الســابقة أن هنــاك طريقتين مألوفتين لتقدير مجموع المجتمع Y للقياسات y_{ij}. فمن أجل عينة عشـــوائية بسيطة من العناقيد يكون المقدر غير المنحاز $y_i = \sum_{i=1}^{M_i} y_{ij} = M_i \,\bar{y}_i$ لمجموع المفـــردة في الوحدة العنقودية i. فإذا فترضنا عينة عشوائية بسيطة حجمها n من الوحدات الــ N في المجتمع. فإن تقديرا غير منحاز لــ Y هو $\hat{Y} = \frac{N}{n} \sum_{i=1}^{n} y_i$ ويكون تباين:

$$V(\hat{Y}) = \frac{N^2(1-f)}{n} \cdot \frac{\sum_{i=1}^{N} (y_i - \bar{Y})^2}{N-1}$$

حيث $\bar{Y} = \frac{Y}{N}$ يمثل متوسط المجتمع على أساس الوحدة العنقودية وغالبا ما نجـــد التقدير $\hat{Y}$ قليل الدقة. ويحدث هذا عندما تختلف الــ $\bar{Y}_i$ (المتوسطات على أســـاس العنصر الواحد) اختلافا بسيطا من وحدة إلى وحدة بينما يتغير M_i كثيرا. وفي هـــذه الحالة يتغير $y_i = M_i \bar{y}_i$ أيضا من وحدة إلى وحدة ويكون التباين $V(\hat{y})$ كبيرا.

ومن أجل عينة عشوائية بسيطة من العناقيد وتقدير النسبة إلى الحجــــم فليكـــن $M_0 = \sum_{i=1}^{N} M_i$ تميل العدد الكلي. وإذا كانت المقادير M_i وبالتالي M_0 كلـــها معروفة، نجد تقديرا بديلا هو التقدير النسبة حيث نتخذ M_i كمتغير مساعد x_i.

(متوسط العينة لكل عنصر) $\hat{Y}_R = M_0 = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} M_i} = M_0$ وفق التقدير النسبة نجد أن

نسبة المجتمع $R = \frac{Y}{X} = \frac{Y}{M_0} = \overline{\overline{Y}}$ وهو متوسط المجتمع لكل عنصر ومن أجل عـــدد العناقيد في العينة كبير نجد

$$V(\hat{Y}_R) = \frac{N^2(1-f)}{n} \cdot \frac{\sum_{i=1}^{N}(y_i - M_i\overline{\overline{Y}})^2}{N-1}$$

$$= \frac{N^2.(1-f)}{n} \cdot \frac{\sum_{i=1}^{N} M_i^2(\overline{Y}_i - \overline{\overline{Y}})^2}{N-1}$$

وهذا يدل على أن تباين $\hat{Y}_R$ يعتمد على التغير بين المتوسطات لكل عنصر درجة أنه على الغالب أصغر من $V(\hat{Y})$.

والتقديرات الموافقة تكون:

$$\hat{\overline{\overline{Y}}} = \frac{\hat{Y}}{M_0} = \frac{N}{nM_0}\sum_{i=1}^{n} y_i \quad , \quad \begin{cases} \hat{\overline{\overline{Y}}}_R = \frac{\hat{Y}_R}{M_0} = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} M_i} \\ = (\text{متوسط العينة لكل عنصر}) \end{cases}$$

وهكذا لا يتطلب $\hat{\overline{\overline{Y}}}_R$ إلا معرفة الـــ M_i التي تقع في العينة التي اخترناها.

2-II-5: المعاينة في حالة احتمال متناسب مع الحجم:

إذا كانت المقادير M_i جميعها معروفة، فتوجد طريقة أخـــرى وهـــي اختيـــار الوحدات باحتمالات متناسبة مع حجومها M_i . وتوضح طريقة اختيـــار وحـــدة

بمفردها من المجتمع الصغير التالي المؤلف من N=7 وحدات وقــد شــكلت المجــاميع التجميعية لـــ M_i. ولاختيار وحدة نسحب عددا عشوائيا بين (1و 30= M_0).

الوحدة	الحجم M_i	$\sum M_i$	المدى المخصص
1	3	3	1-3
2	1	4	4
3	11	15	5-15
4	6	21	16-21
5	4	25	22-25
6	2	27	26-27
7	3	30	28-30

ولنفرض أن هذا العدد 19 ففي العمود $\sum M_i$ يقع العدد 19 في الوحدة الرابعــة التي تغطي المدى من 16 إلى 21 وتغطي الطرفين (16 و 21) ومع هـــذه الطريقــة في السحب يكون احتمال اختيار أي وحدة أخرى متناسبا مع حجمها. ومن أجــل n>1 وافتراض أن المعاينة مع الإعادة. فالاختيار وحدة ثانية وفقا لطريقة التجميع نســحب عددا عشوائيا جديدا بين 1 و 30 وخلافا للمعاينة بدون إعادة لا تحظر اختيار الوحـدة 4 للمرة الثانية. ومع هذه القاعدة، تبقى احتمالات الاختيار متناسبة مع الحجوم عنــد كل سحب وميزة الاختيار مع الإعادة هي بساطة العلاقـــات الخاصــة بالتباينـــات الصحيحة أو المقدرة للتقديرات.

وفي حالة المعاينة بدون إعادة فقد يكون الاحتفاظ باحتمالات اختيار متناسبة مــع الحجوم المختارة أكثر صعوبة ويصبح عاجلا أم آجلا نوعا من المستحيل مع تزايد n.

3-II-5: اختيار باحتمالات غير متساوية مع الإعادة:

لنختبر الوحدة، باحتمال M_i/M_0 ومع الإعادة، حيث $M_0=\sum_{i=1}^{N}M_i$ فسنبيّن أن تقديراً غير منحاز لمجموع المجتمع Y هو

$$\hat{Y}=\frac{M_0}{n}(\bar{y}_1+\bar{y}_2+......+\bar{y}_n)$$

$=M_0$ (متوسط متوسطات الوحدات لكل عنصر)

ونرمز لهذا التقدير بـــ $\hat{Y}_{pPs}$ ويكون أيضاً

$$V(\hat{Y}_{PPS})=\frac{M_0}{n}\sum_{i=1}^{N}M_i(\bar{y}_i-\bar{\bar{Y}})^2$$

بحيث يعتمد تباين $\hat{Y}_{PPs}$ مثله مثل تباين $\hat{Y}_R$ على تغير متوسطات الوحدات لكـــل عنصر. وفي بعض التطبيقات نعلم الحجوم M_i بصورة تقريبية فقط. لذلـــك ســـنعدّ قياس حجم M'_i واحتمال الاختيار $Z_i=\frac{M'_1}{M'_0}$ حيث $M'_0=\sum_{i=1}^{N}M'_i$.

وإلى الحد الذي يتعلق بالنتائج النظرية يمكن أن تكون المقادير Z_i أيّ مجموعة من الأعداد الموجبة مجموعها فوق المجتمع بكامله يساوي الواحد، وسنبيّن أن:

$\hat{Y}_{PPz}=\frac{1}{n}\sum_{i=1}^{n}\frac{y_i}{Z_i}$ هو تقدير غير منحاز لـــ Y بتباين

$$V(\hat{Y}_{PPZ})=\frac{1}{n}\sum_{i=1}^{N}Z_i\left(\frac{y_i}{Z_i}-Y\right)^2$$

فإذا فرضنا أن t_i يمثل عدد المرات يمثل عدد المرات التي تظهر فيها الوحـــدة i في عيّنة محددة حجمها n، حيث يمكن أن يأخذ t_i أيّاً من القيم n ,.... ,1, 2 لنأخذ توزيـــع التكرار المشترك للمقادير t_i من أجل جميع الوحدات الـــــ N في المجتمـــع وطريقـــة سحب عيّنة مكافئة لمسألة الاحتمال المعروفة التي نقذف فيها n كرة إلى N صندوقـــاً،

وعند كل قذفه يمثل Z_i احتمال أن تذهب كرة إلى الصندوق i وبالتالي يكون التوزيع المشترك للمقادير t_i هو توزيع حدودي

$$\frac{n!}{t_1!t_2!.....t_N!}Z_1^{t_1}Z_2^{t_2}.....Z_N^{t_N}$$

وخواص هذا التوزيع هي:

$$E(t_i)=nZ_i \quad ; \; V(t_i)=nZ_i(1-Z_i)$$
$$Cov(t_i t_j)=-nZ_iZ_j \qquad (i\neq j)$$

مبرهنه (1-I-5):

إذا سحبنا عينة n من الوحدات باحتمالات Z_i ومع الإعادة، فعندئذ يكون $\hat{Y}_{PPZ}=\frac{1}{n}\sum_{i=1}^{n}\frac{y_i}{Z_i}$ تقديرا غير منحاز لــ Y بتباين:

$$V(\hat{Y}_{PPZ})=\frac{1}{n}\sum_{i=1}^{n}Z_i\left(\frac{y_i}{Z_i}-Y\right)^2$$

الإثبات:

يمكن كتابة

$$\hat{Y}_{PPZ'}=\frac{1}{n}\left[t_1\frac{y_1}{Z_1}+t_2\frac{y_2}{Z_2}+....+t_N\frac{y_N}{Z_N}\right]=\frac{1}{n}\sum_{i=1}^{N}t_i\frac{y_i}{Z_i}$$

حيث يمتد المجموع فوق جميع الوحدات في المجتمع. وعند تكرار المعاينة تمثل المقادير t المتغيرات العشوائية، بينما الــ y_i والــ Z_i تمثل مجموعة من الأعداد المثبتة ومنه نجد $E(t_i)=nZ_i$ نجد أن:

$$E(\hat{Y}_{PPZ})=\frac{1}{n}\sum_{i=1}^{N}(nZ_i)\frac{y_i}{Z_i}=\sum_{i=1}^{N}y_i=Y$$

وهذا يعني أن $\hat{Y}_{PPZ}$ غير منحاز

وأيضا:

$$V(\hat{y}_{PPZ})=\frac{1}{n^2}\left[\sum_{i=1}^{N}\left(\frac{y_i}{Z_i}\right)^2 V(t_i)+2\sum_{i=1}^{N}\sum_{j>i}^{N}\frac{y_i}{Z_i}\cdot\frac{y_j}{Z_j}Cov(t_i,t_j)\right]$$

$$=\frac{1}{2}\left[\sum_{i=1}^{N}\left(\frac{y_i}{Z_i}\right)^2 Z_i(1-Z_i)-2\sum_{i=1}^{N}\sum_{j>i}^{N}\frac{y_i}{Z_i}\cdot\frac{y_j}{Z_j}Z_i\,Z_j\right]$$

$$=\frac{1}{n}\left(\sum_{i=1}^{N}\frac{y_i^2}{Z_i}-Y^2\right)=\frac{1}{n}\sum_{i=1}^{N}Z_i\left(\frac{y_i}{Z_i}-Y\right)^2$$

وعلى فرض أن $\sum_{i=1}^{n}Z_i=1$. وبأخذ $Z_i=\frac{M_i}{M_0}$ نجد النتـــائج الموافقـــة لمعاينـــة باحتمالات متناسبة مع الحجم.

مبرهنة (2-II-5):

إذا سحبنا، مع الإعادة ، عينة تتضمن n وحدة باحتمالات تتناســب مــع Z_i ، فيكون

$$v(\hat{Y}_{PPZ})=\sum_{i=1}^{n}\frac{\left(\frac{y_i}{Z_i}-\hat{Y}_{PPZ}\right)^2}{n(n-1)}$$

تقديرا غير منحاز لـــ $V(\hat{Y}_{PPZ})$ وذلك من أجل أي n> 1.

الإثبات :

من المطابقة الجبرية المعتادة نجد:

$$\sum_{i=1}^{n}\left(\frac{y_i}{Z_i}-\hat{Y}_{PPZ}\right)^2=\sum_{i=1}^{n}\left(\frac{y_i}{Z_i}-Y\right)^2-n(\hat{Y}_{PPZ}-Y)^2$$

ومن العلاقة الأصلية في المبرهنة:

$$E[n(n-1)v(\hat{y}_{PPZ})]=E\left[\sum_{i=1}^{n}\left(\frac{y_i}{Z_i}-Y\right)^2\right]-nV(\hat{Y}_{PPZ})$$

ومن تعريف $V(\hat{Y}_{PPZ})$ وبإدخال المتغيرات t_i نجد:

$$n(n-1)\,E\left[v(\hat{Y}_{PPZ})\right]=E\left[\sum_{i=1}^{N}t_i\left(\frac{y_i}{Z_i}-Y\right)^2-nV(\hat{Y}_{PPr})\right]$$

$$=n\sum_{i=1}^{n}Z_i\left(\frac{y_i}{Z_i}-Y\right)^2-nV(\hat{Y}_{PPZ})$$

$$=n(n-1)V(\hat{Y}_{PPZ})$$

$$\Rightarrow\ E\left[v(\hat{Y}_{PPZ})\right]=V(\hat{Y}_{PPZ})$$

مبرهنة (3-II-5):

إذا سحبت مع الإعادة، عينة تتضمن n وحدة باحتمالات تتناسب مع الحجـــــم $Z_i=\dfrac{M_i}{M_0}$ فعندئذ يكون.

$$\hat{Y}_{PPZ}=\frac{M_0}{n}\sum_{i=1}^{n}\left(\frac{y_i}{M_i}\right)=\frac{M_0}{n}\sum_{i=1}^{N}(\bar{y}_i)=M_0\bar{\bar{y}}$$

(حيث $\bar{\bar{y}}$ المتوسط غير المرجح المتوسطات الوحدات) تقديرا غير منحاز لــــــ Y بتباين.

$$V(\hat{Y}_{PPs})=\frac{M_0}{n}\sum_{i=1}^{n}M_i(\bar{y}_i-\bar{\bar{Y}})^2$$

وهذه النتائج تتبــــع مــــن المبرهنـــات الســـابقة في هـــذا الفصــــل حيـــث إن

$$\bar{\bar{Y}}=\frac{Y}{M_0}\quad,\quad \bar{y}_i=\frac{y_i}{M_i}$$

مبرهنة (4-II-5):

تحت شروط المبرهنة السابقة يكون

$$v(\hat{Y}_{PPs})=M_0^2\sum_{i=1}^{n}(\bar{y}_i-\bar{y})^2/n(n-1)$$

تقدير عينه غير منحاز لــ $V(\hat{Y}_{PPs})$.

والنتيجة تتبع بتعويض $Z_i = \frac{M_i}{M_0}$ في عبارة V في المبرهنة مثل السابقة وافـــتراض $\bar{y}_i = \frac{y_i}{N_i}$ و $\hat{Y}_{PPZ} = M_0 \bar{\bar{y}}$.

4-II-5: القياس الأمثل للحجم:

في الحالات التي يكون فيها قياس الحجم M_i تقديراً لكبر الوحدة، يبرز ســـؤال مهم هو: ما هو قياس الحجم الذي يجعل تباين $\hat{Y}_{PPZ}$ أصغر ما يمكن:

$$V(\hat{Y}_{PPZ}) = \frac{1}{n}\sum_{i=1}^{N} Z_i \left(\frac{y_i}{Z_i} - y\right)^2$$

$$= \frac{1}{n}\left(\sum_{i=1}^{N} \frac{y_i}{Z_i} - Y^2\right)$$

وتصبح هذه العبارة صفراً إذا كان Z_i يتناسب مع y_i وإذا كانت المقـــادير y_i جميعها موجبة، فهذه المجموعة من القيم Z_i تشكل مجموعة مقبولة من الاحتمـــالات. وبالتالي فإن أفضل قياسات للحجم هي أعداد متناسبة مع مجاميع المفردات y_i في الوحدات المختلفة.

5-II-5: الدقة النسبية لثلاث طرائق:

نقارن هنا دقة الطرائق الثلاث السابقة لتقدير مجموع المجتمع مع وحدات عنقودية ذات حجوم غير متساوية (مفترضين أن M_i معلومة عند الحاجة):

1- اختيار باحتمالات متساوية، تقدير $\hat{Y}_u$

2- اختيار باحتمالات متساوية، تقدير $\hat{Y}_R$

3- اختيار باحتمال متناسب مع الحجم، تقدير $\hat{Y}_{PPS}$

لدراسة الدقة بين الطرائق الثلاث، نلاحظ أن المسألة نعتمد على العلاقـــة بـــين $\bar{y}_i$ و M_i وعلى تباين $\bar{y}_i$ كدالّة في M_i. والحالة الملائمـــة لتقديـــري النســـبة والاحتمال تتناسب مع الحجم هي تلك التي لا يكون $\bar{y}_i$ فيها على صلة بـــــ M_i.

والحالة الملائمة لـــ $\hat{Y}_u$ هي تلك التي لا يكون مجموع الوحدة y_i فيها على صلة بـــ M_i.

لدينا الدراسة السابقة: (بافتراض $\hat{N} \cong (N-1)$) وأن

$E(y_i-\overline{Y})^2=\sum_{i=1}^{N}(y_i-\overline{Y})^2/N$، ويمكن إهمال الانحياز في $\hat{Y}_R$) ففي حالـــة $\hat{Y}_u$ نجد:

$$nV(\hat{Y}_u)=N^2(1-f)E(y_i-\overline{Y})^2=(1-f)E(N\overline{y}_i-\overline{\overline{Y}})^2$$

وفي حالة $\hat{Y}_R$ نجد:

$$nV(\hat{Y}_R)=N^2(1-f)EM_i^2(\overline{y}_i-\overline{Y})^2$$

$$=(1-f)E\left(\frac{M_i}{\overline{M}}\right)^2(M_0\overline{y}_i-Y)^2$$

حيث

$$\overline{M}=\frac{\sum_{i=1}^{N}M_i}{N}=\frac{M_0}{N}$$

ومن أجل $\hat{Y}_{PPS}$ نجد:

$$nV(\hat{Y}_{PPS})=NM_0E\left[M_i(\overline{y}_i-\overline{\overline{Y}})^2\right]=M_0^2E\left[\left(\frac{M_i}{\overline{M}}\right)(\overline{y}_i-\overline{\overline{Y}})^2\right]$$

$$=E\left[\left(\frac{M_i}{\overline{M}}\right)(M_0\overline{y}_i-Y)^2\right]$$

ومن العلاقات الثلاثة السابقة $nV(\hat{Y}_u)$, $nV(\hat{Y}_R)$, $nV(\hat{Y}_{PPS})$ نرى أن $V(\hat{Y}_u)$ يعتمد على دقة المقادير $Ny_i=NM_i\overline{y}_i$ كتقديـــرات لـــــ Y. بينمـــا يعتمـــد $V(\hat{Y}_R)$, $V(\hat{Y}_{PPS})$ على دقة المقادير $M_0\overline{y}_i=M_0\frac{y_i}{M_i}$ كتقديرات لـــــــ Y، وإذا لم يكن $\overline{y}_i$ على صلة بـــ M_i فتوقع أن يكون $M_0\overline{y}_i$ أكثر دقة مـــن الـــــ $NM_i\overline{y}_i$، ونتوقع العكس إذا لم يكن y_i على صلة بـــ M_i.

وبالنسبة لـــ $\hat{Y}_R$, $\hat{Y}_{PPS}$ نلاحظ من $nv(\hat{Y}_R')$, $nV(\hat{Y}_{PPS})$ أن $V(\hat{Y}_R)$ يعطـــــي للوحدات الكبيرة ترجيحه أكبر نسبيا مما يعطيـــــه $V(\hat{Y}_{PPS})$. ونلاحـــظ أيضـــا أن $\hat{Y}_u$, $\hat{Y}_R$ يستفيدان من حد التصحيح الذي يمكن أن يصبح كبيرا في طبقات صغـــيرة مثلا $N_h=10$, $n_h=2$ وقد حدث هذه النقطة بتطوير الاختيار باحتمـــالات غـــير متساوية بدون إعادة. وتصبح العلاقة في $nV(\hat{Y}_R)$ بالطبع في العينات الكبيرة فقط.

6-II-5: المعاينة باحتمالات غير متساوية دون إعادة:

في دراسة إحصائية واسعة قسمت فيها الوحدات العنقودية أولا، وفق (مثلا الموقع الجغرافي) إلى عدد كبير من الطبقات الصغيرة نسبيا، ثم سحب عدد صغير فقط مـــن الوحدات العنقودية من كل طبقة، والحالة $n_h=2$، التي تقدم درجة حرية واحدة من كل طبقة لتقدير أخطاء المعاينة، هي حالة ذات أهمية خاصة. لنفترض أننـــا ســـحبنا وحدتين من طبقة، وأن الوحدة الأولى مسحوبة باحتمالات Z_j، تتناسب مع قيـــاس ما للحجم ولتكن الوحدة i هي الوحدة المختارة. فإذا اتبعنا الطريقة الأكثر بداهـــة، فإننا نختار في السحب الثاني إحدى الوحدات الباقية بعد تخصيص احتمالات هي:

$\frac{Z_j}{1-Z_i}$ وبالتالي يكون الاحتمال الكلي لاختيار الوحدة i في أي من الســـحبين الأول أو الثاني هو

$$\pi_i = Z_i + \sum_{i\neq j}^{N} \frac{Z_j Z_i}{(1-Z_j)} = Z_i\left(1+\sum_{j\neq i}^{N}\frac{Z_j}{1-Z_j}\right)$$

$$= Z_i\left(1+A-\frac{Z_i}{1-Z_i}\right)$$

حيث $A=\sum_{i=1}^{N}\frac{Z_j}{1-Z_j}$ مأخوذ من الوحدات الـــ N جميعها.

لنفترض أن $\pi_i = 2Z_i$ فالاحتمالات النسبية لاختيار الوحدات الباقية متناسبة مــع قياس للحجم هو Z_i. ومقدر العينة لـــ Hervitz–Thompson هو:

$$\hat{Y}_{HT} = \sum_{i=1}^{2} \frac{y_i}{\pi_i} = \frac{1}{2} \sum_{i=1}^{2} \frac{y_i}{Z_i}$$

وسيكون له تباين معدوم في حالة تناسب Z_i مع y_i حيث $Z_i = \frac{y_i}{Y}$، وستعطي كل وحدة اختيرت التقدير الصحيح $\frac{y_i}{z_i} = Y$. إلا أن المقادير $Z_i' = \frac{\pi_i}{z}$ في العلاقة π_i الأخيرة ستكون دائما أقرب إلى التساوي من المقادير Z_i الأصلية بسبب العامل الثاني فيها.

وسنقدم فيما يلي التقدير العام الأكثر شهرة لمجموع مجتمــــع في حالــــة معاينـــة باحتمالات غير متساوية وبدون إعادة.

7-II-5: مقدر هيرفتز تومبسون:

اختيرت عينة من n وحدة، بدون إعادة، وفق طريقة ما وليكن π_i احتمال أنــــه تكون الوحدة i ضمن العينة ، π_{ij} احتمال أن تكون الوحدتان i وj كلتاهما ضمـــــن العينة فتصبح عندئذ العلاقة التالية:

$$\sum_{i=1}^{N} \pi_i = n \,;\, \sum_{j \neq i}^{N} \pi_{ij} = (n-1)\pi_i \,,\, \sum_{i=1}^{N} \sum_{j>i} \pi_{ij} = \frac{n(n-1)}{2}$$

فإذا رمزنا بـــ P(s) لاحتمال أن تتضمن عينة n من الوحدات المحــــددة فعندئـــذ يكون $\pi_{ij} = \sum P(s)$ والمجموع على كل العينات التي تتضمن الوحدتين الـــ i والـــــ j، $\pi_i = \sum P(s)$ والمجموع على كل العينات التي تتضمن الوحدة i. وعندما نــــأخذ $\sum \pi_{ij}$ حيث $j \neq i$ ، فإن كل $P(S)$ لعينة تحوي الوحــــدة i يتكـــرر تعدادهـــا في المجموع (n-1) مرة، ذلك لأن العينة تتضمن (n-1) من القيم الممكنة لـــ j وهذا يثبـــت

العلاقة الثانية. أما العلاقة الثالثة فينتج من الثانية ومقدر HT لمجمـــوع المجتمـــع هـــو $\hat{Y}_{HT}=\sum_{i=1}^{n}\frac{y_i}{\pi_i}$ حيث y_i القياس من الوحدة i.

مبرهنة (5-II-5) : (See: P260:S.T. Cochran)

إذا كان $\pi_i>0$ ، (i = 1, 2, ..., N) فعندئذ يكون $\hat{Y}_{HT}=\sum_{i=1}^{n}\frac{y_i}{\pi_i}$ تقديرا غير منحاز لـــ Y يتباين:

$$V(\hat{Y}_{HT})=\sum_{i=1}^{N}\frac{(1-\pi_i)}{\pi_i}y_i^2+2\sum_{i=1}^{N}\sum_{j>i}^{N}\frac{(\pi_{ij}-\pi_i\pi_j)}{\pi_i\pi_j}y_iy_j$$

حيث π_{ij} تمثل احتمال أن تكون الوحدتان i ، و j ضمن العينة.

5-II-8: طريقة مورفي : هنا تسحب الوحدات المتتالية باحتمالات

$$Z_i=\frac{Z_j}{1-Z_i},\frac{Z_K}{1-Z_i-Z_j},...$$

ويتبع مقدر مورفي بأنه في أي تقدير مرتب من هذا الفصل من التقديرات يمكـــن إقامة تقدير غير مرتب وغير منحاز أيضا وله تباين أصغر. ومقدرة المقترح هو

$$\hat{Y}_M=\frac{\sum_{i=1}^{N}P\left(S/_i\right)y_i}{P(s)}$$ حيث:

$P\left(S/_i\right)$ يمثل الاحتمال الشرطي للحصول على مجموعة الوحدات المسحوبة، علما بأن الوحدة الـــ i قد سحبت أولا.

$P(s)$ يمثل الاحتمال غير الشرطين للحصول على مجموعة الوحدة المسحوبة وإن $\hat{Y}_M$ هو مقدر غير منحاز لـــ Y بتباين :

$$V(\hat{Y}_M)=\sum_{i=1}^{N}\sum_{j>i}^{N}\frac{Z_iZ_j(1-Z_i-Z_j)}{2-Z_i-Z_j}\cdot\left(\frac{y_i}{Z_i}-\frac{y_j}{Z_j}\right)^2$$

حيث نجد أن $V(\hat{Y}_M)<V(\hat{Y}_{PPZ})$.

9-II-5 : تمارين غير محلولة:

تمرين (1):

من أجل المعلومات الإحصائية الواردة في المثال (5-1): قارن الدقة النسبية الصافية للأنواع الأربعة من الوحدات، وذلك عندما يكون الهدف هو تقدير العدد الكلي للأغراس في المسكبه بخطأ معياري يساوي 200 غرس (لاحظ أن عامل التصحيحي مأخوذ بالحسبان).

تمرين (2):

قسمنا مجتمعا من 2500 من العناصر إلى عشر طبقات، تحوي كل منها 5 وحدات كبرى تتألف كل منها من 50 عنصرا، وتحليل التباين للمجتمع، على أساس العنصر الواحد، ولمفردة واحدة، هو كما يلي:

مصدر التغير	درجة الحرية	متوسط للمربعات
ما بين الطبقات	9	30.6
مابين الوحدات الكبرى ضمن الطبقات	490	3.0
ما بين العناصر ضمن الوحدات الكبرى	2000	1.6

متجاهلين عامل التصحيح، هل تكون الدقة النسبية للوحدة الكبيرة إلى الوحدة الصغيرة أكبر في حالة المعاينة العشوائية البسيطة منها في المعاينة العشوائية الطبقية (محاصلة تتناسبية)؟

تمرين (3):

في دراسة إحصائية في الريف كانت وحدة المعاينة عنقودا من المزارع، ووجدنا أن كلفة أخذ عينة تتضمن n وحده هي:

$$C=4tMn+60\sqrt{n}$$

حيث t هو الزمن بالساعات الذي نقطيه في الحصول على الأجوبة من مـــزارع واحد. وإذا أنفقنا 2000 دولار على هذه الدراسة. نجد أن قيم n في حالة M= 1, 2, 10، $t=\frac{1}{2},2$ هي كما يلي:

	M		
	1	5	10
$t=\frac{1}{2}$	400	131	74
$t=2$	156	40	21

تحقق من قيمتين من هذه القيم للتأكد من أنك تفهم استخدام العلاقـــة وتبـــاين متوسط العينة مع تجاهل معامل التصحيح هو :

$\frac{S^2}{Mn}[1+(M-1)\rho]$ وإذا كان $\rho=0.1$ من أجل جميع قيم M يبـــين 1 و 10.

فما هو حجم الوحدة الأكثر دقة في حالة:

a- $t=\frac{1}{2}$ ساعة

b- $t=2$ ساعة.

كيف تفسر الفرق بين النتيجتين؟

إذا توفر 5000 دولار للدراسة الإحصائية، فهل نتوقع تنـــاقص الحجـــم الأمثـــل للوحدة (بالمقارنة مع 2000 دولار) أم تزايده؟ فما هي الأسباب.

أوجد الحجم الأمثل للتحقق من صحة المحاكمة؟.

تمرين (4):

يعطي هورفيز وتوميسون البيان الإحصائي التالي لتقديرات بالعين المجـــردة M_i وللأعداد الفعلية y_i للمنازل في جادات مدينة آميـــس في أيـــووا. وللمســـاعدة في

الحسابات، أعطيت أيضاً قيم. $\bar{y}_i$ و $\frac{\bar{y}_i^2}{M_i}$. وقد اختيرت عينّة من جادة واحـــــدة.

احسب تباينات العدد الكلي للمنازل Y، كما نحصل عليه من:

a- التقدير غير المنحاز في معاينة باحتمالات متساوية.

b- التقدير النسبة باحتمالات متساوية.

C- معاينة باحتمال متناسب مع M_i.

وهل تتفق النتائج ما رأيناه في الدراسة النظرية.

M_i	y_i	$\bar{y}_i$	y_i^2/M_i
9	9	1.0000	9.000
9	13	1.4444	18.778
12	12	1.0000	12.000
12	12	1.0000	12.000
12	14	1.1667	16.333
14	17	1.2143	20.643
14	15	1.0714	16.071
17	20	1.1765	23.529
18	19	1.0556	20.056
18	18	1.0000	18.000
19	19	1.000	19.000
21	25	1.1905	29.762
23	27	1.1739	31.696
24	21	0.8750	18.375
24	35	0.4583	51.042
25	22	0.8800	19.360
26	25	0.9615	24.038
27	27	1.000	27.000
30	47	1.5667	73.633

40	37	0.9250	34.225

تمرين (5) :

إذا كان N= 3 في مجتمع $\frac{1}{6}, \frac{1}{3}, \frac{1}{2} = Z_i$ و $y_i = 7,5,2$ وسحبنا وحدتين بدون إعادة ، الأولى باحتمال متناسب مع z_i والثانية باحتمال متناسب مع الحجوم الباقية.

a- تحقق من أن :

$$\pi_1 = \frac{51}{60} ; \pi_2 = \frac{44}{60} ; \pi_3 = \frac{25}{60} ; \pi_{12} = \frac{35}{60}$$

$$\pi_{13} = \frac{16}{60} ; \pi_{23} = \frac{9}{60} ;$$

b- من أجل هذه الطريقة في اختبار العينة قارن تباين $\hat{Y}_{HT}$ و $\hat{Y}_M$

الفصل السادس
تطبيق المعاينة العشوائية في نظرية الاستقراء الإحصائي

6-1: مقدمة:

إن نظرية الاستقراء الإحصائي والتي أصبحت تعرف حديثاً بنظرية القرار هي تلك الطرائق التي تمكنّ الإحصائي من إجراء تعميمات حول المجتمع الإحصائي بدءاً من معطيات العينة المختارة منه. وسنعالج هنا كيفية تقدير وسطاء المجتمع (متوسط، تباين، انحراف معياري..) وذلك من معرفة متوسط وتباين والانحراف المعياري للعينة المسحوبة منه والتي ندعوها بإحصائيات. فقد تكون الدراسة المطروحة هي اختيار طريقة في الإنتاج وأفضليتها على أخرى أو اختيار مصل جديد في معالجة مرض معين أو اختيار فعالية سماد معين في نمو نبات معين ومقارنته بأنواع أخرى من الأسمدة. فالاستقراء الإحصائي هو الانتقال من الجزء إلى الكل في ضوء التخمين والتقدير، وإن تقدير وسيط مجتمع إحصائي ما، يمكن أن يعطى بشكل نقطي أو بشكل مجالي، مثلاً متوسط العينة $\overline{X}$ والمسحوب من عينة حجمها n هو تقدير لمتوسط المجتمع μ وكذلك نسبة العينة $\hat{P}=\frac{x}{n}$ (حيث x عدد النجاحات و n عدد التكرارات المستقلة لتجربة برنولية) هو تقدير لنسبة المجتمع p في توزيع داني.

ولكن ما من سبب يدعونا إلى التوقع أن التقدير النقطي يكون مساوٍ تماماً لوسيط المجتمع، لذلك من الأفضل أن تنشئ مجالاً نتوقع أن نجد فيه قيمة الوسيط، ومثل هذا المجال ندعوه بالتقدير المجالي ، أما الأول فندعوه بالتقدير النقطي. وللحصول على التقدير المجالي للوسيط μ منشيء مجالاً من الشكل $\overline{X}\pm K$ حيث K تتعين من التوزيع العيني لـ $\overline{X}$. والآن بدلاً من إدعاء أن $\overline{X}$ مساوٍ لـ μ بالضبط، فإننا نشعر بثقة أكبر بكتابه أن

$$\mu \in [\overline{X}-K,\ \overline{X}+K]$$

وعلينا أن نحدد هنا القاعدة التي تمكننا من حساب الوسيط وندعوهـــا بـــالمقدر وقيمة الوسيط نفسه وندعوها بالتقدير وعلينا أن نبحث عن مقدر لا يختلف كثيرا عن قيمة الوسيط نفسه (القيمة الحقيقية) وبالتالي يفضل جدا أن نتعامل مع مقدرات غـــير منحازة ومتماسكة وفعالة كما شرحنا ذلك في الفصل الأول.

2-6-: تقدير متوسط (مجتمع إحصائي) μ تباينة (σ^2) معلوم:

ليكن لدينا مجتمع إحصائي توقعه μ وتباينه σ^2 معلـــوم (أو يمكـــن تقديـــره من S^2). والمطلوب بناء مجال ثقة حول μ بثقة $(1-\alpha)$. فالإنجاز ذلك نسحب عينة عشوائية من الحجم n من ذلك المجتمع ونرمز لمقادير العينة $x_1,\ x_2,\ ....,x_n$ ولمتوســط العينة $\overline{X}$.

إن $\overline{X}$ يتوزع طبيعا بتوقع μ وتباين $\frac{\sigma^2}{n}$ وذلك إذا كان المجتمع طبيعيا أو كان $n \geq 30$ مستندين إلى مبرهنه النهاية المركزية ومنه

$$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \Rightarrow Z = \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

ومن العلاقة الاحتمالية في التوزيع الطبيعي المعياري:

$$P\left[-Z_{\alpha/2} < Z < Z_{\alpha/2}\right] = 1-\alpha$$

وبتعويض Z بما يساويها نجد:

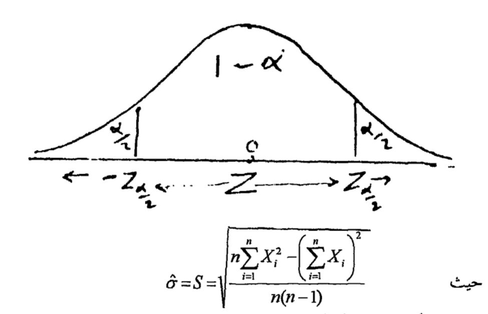
$$P\left[-Z_{\frac{\alpha}{2}} < \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < Z_{\frac{\alpha}{2}}\right] = 1-\alpha$$

$$\Leftrightarrow P\left[\overline{X} - Z_{\alpha/2}.\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + Z_{\alpha/2}.\frac{\sigma}{\sqrt{n}}\right] = 1-\alpha$$

وبالتالي نكون واثقين بمقدار $(1-\alpha)$ من أن

$$\overline{X} - Z_{\alpha/2}.\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + Z_{\alpha/2}.\frac{\sigma}{\sqrt{n}}$$

وهو ما ندعوه بـــ $(1-\alpha)$ مجال ثقه حول μ.



حيث
$$\hat{\sigma}=S=\sqrt{\frac{n\sum_{i=1}^{n}X_i^2-\left(\sum_{i=1}^{n}X_i\right)^2}{n(n-1)}}$$

و $Z_{\alpha/2}$ تعيّن من جدول التوزيع الطبيعي المعياري حيث نستعرض فيما يلي بعض القيم الشهيرة من أجل مستوى ثقه معين.

$1-\alpha$	0.90	0.95	0.98	0.99
$Z_{\alpha/2}$	1.645	0.96	2.33	2.58

نسمي المقدار $e=\left|Z_{\alpha/2}.\frac{\sigma}{\sqrt{n}}\right|$ بالحد الأعلى لخطأ التقدير. ويمكننا من أجل دقة معينة أن نحدّد حجم العينة المناسب وذلك من استخدام علاقة الخطأ السابقة حيث نجد:

$$n=\left[\frac{Z_{\alpha/2}.\sigma}{e}\right]^2$$

حيث n تكون حجم العينة المناسب لتقدير μ بوساطة $\overline{X}$ وبثقة $(1-\alpha)$ وبخطأ أعظمي لا يتجاوز e.

مثال (6-1):

يمثل البيان الإحصائي التالي إنتاج عشر عزسات من الطماطم مقيسا بالكغ.

فإذا علمت أن قياسات الإنتـــاج في 2.3, 2.6, 2.2, 3.1, 4.0, 1.9, 2.7, 1.9, 3.3, 3.0 مجتمع عزسات الطماطم توصف بتوزيع طبيعي يتباين $\sigma^2 = 0.36$ فالمطلوب:

1- احسب مقدار الخطأ الأعظمي المرتكب لتقدير μ مجالياً وذلك بثقة 0.99, 0.95, 0.90.

2- عين 0.99, 0.95, 0.90 مجال ثقه حول متوسط الإنتاج μ.

3- عيّن حجم العينة المناسب لتقدير متوسط إنتاج عزسة الطماطم μ بحيـــث لا يزيد الخطأ على 0.2 كغ إلا باحتمال زهيد لا يتجاوز 0.01.

الحل: نحسب أولاً متوسط العينة:

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{2.3 + 2.6 + ... + 3.0}{10} = 2.7 \qquad (1)$$

ويكون الخطأ الأعظمي المرتكب بثقة 0.90:

$$e = \left| \frac{\pm Z_{\alpha/2} . \sigma}{\sqrt{n}} \right| = \left| \frac{\pm (1.645) . \sqrt{0.36}}{\sqrt{10}} \right| = 0.3$$

والخطأ الأعظمي المرتكب بثقة 0.95:

$$e = \left| \pm (0.96) . \frac{\sqrt{0.36}}{\sqrt{10}} \right| = 0.37$$

والخطأ الأعظمي المرتكب بثقة 0.99:

$$e = \left| \pm (2.58) . \frac{\sqrt{0.36}}{\sqrt{10}} \right| = 0.49$$

(2) إن $(1-\alpha)$ مجال ثقة حول μ هو

$$\overline{X} - e < \mu < \overline{X} + e$$

فمن أجل $1-\alpha = 0.90$ يكون

$$2.7 - 0.31 < \mu < 2.7 + 0.31$$

$$2.39 < \mu < 3.01$$

أي أن متوسط إنتاج عزسه الطماطم وبثقة 90% لن يقل عن 2.39 كغ ولن يزيد عن 3.01 كغ.

ومن أجل $1-\alpha=0.95$ يكون:

$$2.7-0.37<\mu<2.7+0.37$$

$$2.33<\mu<3.07$$

ومن أجل $1-\alpha=0.99$ يكون:

$$2.7-0.49<\mu<2.7+0.49$$

$$2.21<\mu<3.19$$

ونلاحظ هنا أن مجال الثقة يتسع مع ازدياد أمثال الثقة.

(3) إن حجم العينة يتحدد من العلاقة:

$$n=\left[\frac{Z_{\alpha/2}.\sigma}{e}\right]^2=\left[\frac{(2.58)(\sqrt{0.36})}{(0.2)}\right]^2\approx 59.9$$

أي أن حجم العينة يجب أن لا يقل عن 60 غرسة.

3-6: التقدير المجالي لمتوسط مجتمع طبيعي بعيّنات صغيرة الحجم:

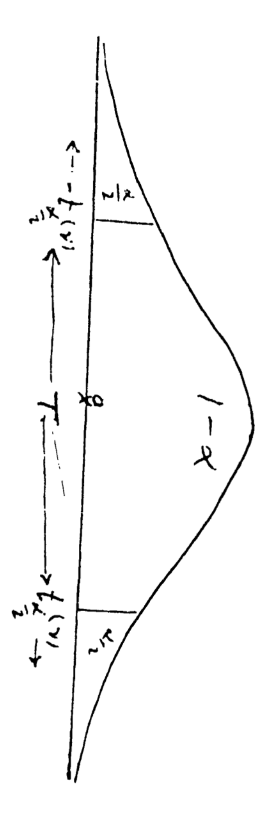
من أجل مجتمع إحصائي يتوزع على وجه التقريب طبيعياً يتوقع μ مجهول وتباين σ^2 معلوم أو يمكن تقديره من تباين العينة المسحوبة منه S^2. والمطلوب بناء مجال ثقة حول μ بمستوى ثقة $(1-\alpha)$. ولإنجاز ذلك نسحب عيّنة عشوائية من ذلك المجتمع من الحجم n أقل من 30 ونستخدم الإحصاء

$$T=\frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}}$$

والذي يتوزع وفق ستودنت بدرجة من الحرية $\nu=n-1$ و $\hat{\sigma}=S$ (كما رأينا في مقرر مبادئ الإحصاء النظري) حيث إن توزيع ستودنت يشبه تقريباً التوزيع الطبيعي المعياري لكنه أكثر تفلطحاً.

ونستخدم العلاقة الاحتمالية في توزيع ستودنت:

$$P\left[-t_{\alpha/2}(\nu)<T<t_{\alpha/2}(\nu)\right]=1-\alpha$$


α/2
1 − α
α/2

وباستبدال T بما يساويها:

$$P\left[-t_{\alpha/2}(\nu)<\frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}}<t_{\alpha/2}(\nu)\right]=1-\alpha$$

$$\Rightarrow P\left[\bar{X}-t_{\alpha/2}(\nu)<\frac{s}{\sqrt{n}}<\mu<\bar{x}+t_{\alpha/2}(\nu).\frac{s}{\sqrt{n}}\right]=1-\alpha$$

أي نكون واثقين بمقدار $(1-\alpha)$ من أن:

$$\mu\in\left[\bar{X}-t_{\alpha/2}(\nu)<\frac{s}{\sqrt{n}}<\mu\ .\bar{X}+t_{\alpha/2}(\nu).\frac{s}{\sqrt{n}}\right]$$

والمجال الأخير ندعوه بمجال ثقة حول μ وبثقة $(1-\alpha)$ في حالة عيّنات صغيرة الحجم ومجتمعات تتوزع على وجه التقريب توزعاً طبيعياً. حيث $t_{\alpha/2}(\nu)$ تتعين من جدول توزيع ستودنت وفقاً لدرجة الحرية $(\nu=n-1)$ والاحتمال $(1-\alpha)$.

و $\bar{X}$ متوسط العينة المسحوبة وs الانحراف المعياري للعينة وهو معطى في الفقرة (2-6) ونسمي المقدار $e=\left|t_{\alpha/2}(\nu)\frac{s}{\sqrt{n}}\right|$ بالخطأ الأعظمي المرتكب في تقدير μ بوساطة $\bar{X}$ وبثقة $(1-\alpha)$.

ويصبح مجال الثقة من الشكل $\bar{X}-e<\mu<\bar{X}+e$ ويمكن تقدير حجم العينة المناسبة لتقدير μ بوساطة $\bar{X}$ وبثقة $(1-\alpha)$ وبخطأ لا يتجاوز e وذلك من العلاقة في e السابقة كما يلي:

$$n=\left[\frac{t_{\alpha/2}.S}{e}\right]^2$$

مثال (6-2):

قمنا بقياس ارتفاع 15 غرسة من نبات الباذنجان بعد فترة من زرعها، فكان متوسط الارتفاع 83 c.m وبانحراف معياري 5.8 c.m عين 95% مجال ثقة حول

متوسط الارتفاع في حقل الباذنجان الذي اخترنا منه الغرسات مع العلم بأن ارتفـــاع الغرسات يتوزع على وجه التقريب طبيعيا.

الحل:

إن $(1-\alpha)$ مجال ثقة حول μ هو

$$\overline{X}-e<\mu<\overline{X}+e$$

حيث من أجل $1-\alpha=0.95$ نجد أن $\alpha=0.05$

و $\frac{\alpha}{2}=0.025$, $\nu=n-1=15-1=14$

وبالتالي من جدول ستودنت نجد قيمة:

$$t_{\alpha/2}(\nu)=t_{0.25}(14)=2.145$$

والخطأ المرتكب

$$e=\left|t_{\alpha/2}(\nu).\frac{s}{\sqrt{n}}\right|=\left|(2.145).\frac{(5.8)}{\sqrt{15}}\right|=3.2$$

ومنه 95% مجال ثقة حول μ يكون:

$$83-3.21<\mu<83+3.21$$

$$79.79<\mu<86.21$$

أي نكون واثقين بمقدار 95% من أن متوسط ارتفاع الغرسة في الحقل لن يقل عن c.m 79.79 ولن يزيد على 86.21 c.m.

مثال (3-6)

خضعت عينة من 12 فأرا تجريبيا لنظام غذائي معين خلال الأشهر الثلاثـــة الأولى من حياتها وقيست الزيادة في الوزن بالغرام لكل فأر فكانت كما يلي:

55, 62, 54, 58, 65, 64, 60, 62, 59, 67, 62, 61

والمطلوب:

1- عين 90% مجال ثقة لمتوسط الزيادة في الوزن وفق التجربة الســـابقة في حيـــاة مجتمع الفئران علما بأن الزيادة في الوزن تخضع على وجه التقريب للتوزيع الطبيعي.

2- ما هو حجم العينة المناسب لتقدير μ (متوسط الزيادة في الوزن) بثقة 0.90 وبخطأ لا يتجاوز غراما واحدا.

الحل:

1- إن $1-\alpha=0.90$ مجال ثقة حول μ (متوسط الزيادة في الوزن) هو من الشكل:

$$\overline{X}-t_{\alpha/2}.\frac{s}{\sqrt{n}}<\mu<\overline{X}+t_{\alpha/2}.\frac{s}{\sqrt{n}}$$

حيث

$$\overline{X}=\frac{\sum_{i=1}^{n}X_i}{N}=\frac{55+62+....+61}{12}=60.75$$

$$S=\sqrt{\frac{n\sum_{i=1}^{n}X_i^2-\left(\sum_{i=1}^{n}x_i\right)^2}{n(n-1)}}=\sqrt{\frac{(12)(44449)-(729)^2}{(12)(11)}}$$

$$=3.84$$

$$1-\alpha=0.90\Rightarrow\alpha=0.10\Rightarrow\frac{\alpha}{2}=0.05$$

$$v=n-1=12-1=11$$

ومن جدول ستودنت

$$t_{\alpha/2}(v)=t_{0.05}(11)=1.796$$

وبالتالي إن 90% مجال ثقة حول μ هو من الشكل :

$$60.75-(1.796).\frac{3.84}{\sqrt{12}}<\mu<60.75+(10796).\frac{3.84}{\sqrt{12}}$$

$$60.75-1.991<\mu<60.75+1.991$$

$$58.759<\mu<62,741$$

2- إن حجم العينة المطلوب يحسب من العلاقة

$$n=\left[\frac{t_{\alpha/2}.S}{e}\right]^2=\left[\frac{(1.796)(3.84)}{1}\right]^2\approx48$$

أي أن حجم العينة يجب أن لا يقل عن 48 لكي تقدر μ بوساطة X بثقة 0.90 وخطأ لا يتجاوز غراما واحدا.

4-6: التقدير المجالي لنسبة:

لنفرض أن صنفا معينا A يوجد في مجتمع كبير بنسبة تساوي P. وسحبنا عينة عشوائية حجمها n من هذا المجتمع، وعرفنا النجاح بأنه الحصول على عنصر من الصنف A، فإن احتمال النجاح عند كل سحب هو عمليا P، ونسبة النجاح في العينة هي عدد عناصر الصنف A ولنرمز لها بـ X (عدد النجاحات) مقسوما على حجم العينة n أي $\hat{P}=\frac{X}{n}$ وبناء على فكرة تقريب التوزيع الثنائي بالتوزيع الطبيعي وجدنا أنه يمكن اقتراض أن عدد النجاحات X (مجموع عينة) و $\hat{P}$ هو متوسط العينة وبتطبيق مبرهنة النهاية المركزية على X يسمح لنا القول إن

$$X \sim N(nP, nPQ) \text{ (تقريبا) } \Leftrightarrow$$

$$\hat{P} \sim N\left(P, \frac{PQ}{n}\right) \quad , \quad Q=1-P$$

حيث

$$E(\hat{P})=E\left(\frac{X}{n}\right)=\frac{E(X)}{n}=\frac{nP}{n}=P$$

$$V(\hat{P})=V\left(\frac{X}{n}\right)=\frac{V(X)}{n^2}=\frac{nPQ}{n^2}=\frac{PQ}{n}$$

وهذا كله تحت شرط أن $n \geq 30$ وبالتالي :

$$Z=\frac{\hat{P}-P}{\sqrt{\frac{PQ}{n}}} \sim N(0,1)$$

ومنه باستخدام العلاقة الاحتمالية للتوزيع الطبيعي الميعاري:

$$P\left[-Z_{\alpha/2} < Z < Z_{\alpha/2}\right]=1-\alpha$$

نجد أن :

$$P\left[-Z_{\alpha/2}<\frac{\hat{P}-P}{\sqrt{\frac{PQ}{n}}}<Z_{\alpha/2}\right]=1-\alpha$$

$$\Rightarrow P\left[\hat{P}-Z_{\alpha/2}.\sqrt{\frac{PQ}{n}}<P<\hat{P}+Z_{\alpha/2}.\sqrt{\frac{PQ}{n}}\right]=1-\alpha$$

أي نكون واثقين بمقدار $(1-\alpha)$ من أن :

$$\hat{P}-Z_{\alpha/2}\sqrt{\frac{PQ}{n}}<P<\hat{P}+Z_{\alpha/2}.\sqrt{\frac{PQ}{n}}$$

ولكن المقدار $\sqrt{\frac{PQ}{n}}$ يحوي مجاهيل. مما يمنع اعتماد هذا المجال و(أن الانحـــراف المعياري يتغير ببطيء شديد عند تغير n في التوزيع الحداني (الثنائي) عندئـــذ يمكننـــا الاستعاضة عن

$\sqrt{\frac{PQ}{n}}$ بـ $\sqrt{\frac{\hat{P}\hat{q}}{n}}$ حيث $\hat{q}=1-\hat{P}$

وبالتالي يصبح $(1-\alpha)$ مجال ثقة حول P من الشكل

$$\hat{P}-Z_{\alpha/2}\sqrt{\frac{\hat{\mu}\hat{q}}{n}}<P<\hat{P}+Z_{\alpha/2}.\sqrt{\frac{\hat{P}\hat{q}}{n}}$$

ونسمي المقدار

$e=\left|\pm Z_{\alpha/2}.\sqrt{\frac{\hat{P}\hat{q}}{n}}\right|$ بالخطأ الأعظمي المرتكب لتقديـــر P بوســـاطة $\hat{P}$ وبثقـــة $(1-\alpha)$ ويمكننا حساب حجم العينة المناسب لتقدير P بوساطة $\hat{P}$ وبثقـــة $(1-\alpha)$ وبخطأ لا يتجاوز e من عبارة e السابقة كما يلي:

$$n=\frac{Z^2_{\alpha/2}.\hat{P}.\hat{q}}{e^2}$$

مثال (4-6):

تضمنت عينة من 250 من طلبة الجامعة 30 طالبا يستخدم اليد اليسرى في الكتابة.

1- حدد 95% مجال ثقة حول النسبة الحقيقية P للطلاب اليسراويين في الجامعة.

2- ما هو حجم العينة اللازم لتقدير P بوساطة $\hat{P}$ وبثقة 0.99 وبخطأ لا يتجاوز 0.01.

الحل:

1- إن $\hat{P}=\frac{X}{n}=\frac{30}{250}=0.12$ وإن 0.95 مجال ثقة حول النسبة الحقيقية P للطلاب اليسراويين في الجامعة يكون من الشكل :

$$P-Z_{\alpha/2}\sqrt{\frac{\hat{P}\hat{q}}{n}}<P<\hat{P}+Z_{\alpha/2}\sqrt{\frac{\hat{P}\hat{q}}{n}}$$

$$0.12-(1.96).\sqrt{\frac{(0.12)(0.88)}{250}}<P<0.12+1.96\sqrt{\frac{(0.12)(0.88)}{250}}$$

$$0.12-0.04<P<0.12+0.04$$

$$0.08<P<0.16$$

أي تكون واثقين 95% من أن نسبة المستخدمين لليد اليسرى في الكتابة في الجامعة لن تقل عن 0.08 ولن تزيد على 0.16.

-2 $$n=\left[\frac{Z^2_{\alpha/2}.\hat{P}.\hat{q}}{e^2}\right]=\left[\frac{(2.58)^2(0.12)(0.88)}{(0.01)^2}\right]\approx 7029$$

أي يجب أن لا يقل حجم العينة عن 7029 من أجل تقدير P بوساطة $\hat{P}$ بثقة 99% وبخطأ لا يتجاوز 0.01.

6-5: التقدير المجالي للفرق بين متوسطي مجتمعين إحصائين بعينات كبيرة الحجم:

في كثير من الأحيان نرغب في مقارنة متوسطي مجتمعين، كمقارنة متوسطي الدخل في بلدين، أو مقارنة طريقتين في التعليم، أو مقارنة جودة تصنيع أجهزة طبية أو إلكترونية من نوعين مختلفين، فمن أجل ذلك، نفترض أنه لدينا مجتمعين إحصائيين توقع الأول μ_1 وتباين σ_1^2 وتوقع الثاني μ_2 وتباين σ_2^2، نسحب من المجتمع الأول

عينة من الحجم n_1 ونسحب من المجتمع الثاني عينة من الحجم n_2 وبشكل مستقل عن الأولى.

وليكن $\overline{X}_1$ و S_1^2 متوسط وتباين العينة الأولى. $S_2^2, \overline{X}_2$ متوسط وتباين العينة الثانية.

والمطلوب بناء $(1-\alpha)$ مجال ثقة حول $(\mu_1-\mu_2)$ مع العلم أن

$$\hat{\sigma}_2^2=S_2^2 \quad , \quad \hat{\sigma}_1^2=S_1^2$$

وفي حالة كون المجتمعات طبيعية أو كانت $n_1 \ ,n_2 \geq 30$ (وبتطبيق مبرهنة النهاية المركزية يكون:

$$\left.\begin{array}{l}\overline{X}_1 \sim N\left(\mu_1,\dfrac{\sigma_1^2}{n_1}\right)\\ \overline{X}_2 \sim N\left(\mu_2,\dfrac{\sigma_2^2}{n_2}\right)\end{array}\right\} \Rightarrow \overline{X}_1-X_2 \sim N\left(\mu_1-\mu_2,\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}\right)$$

ومنه

$$Z=\frac{(X_1-X_2)-(\mu_1-\mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}} \sim N(0.1)$$

$$\Leftrightarrow \ Z=\frac{(\overline{X}_1-\overline{X}_2)-(\mu_1-\mu_2)}{\sqrt{\dfrac{S_1^2}{n_1}+\dfrac{S_2^2}{n_2}}} \sim N(0.1)$$

وحسب (2.6) باستخدام:

$$P\left[-Z_{\alpha/2}<Z<Z_{\alpha/2}\right]=1-\alpha$$

نجد أن:

$$P=\left[-Z_{\alpha/2}<\frac{(\overline{X}_1-\overline{X}_2)-(\mu_1-\mu_2)}{\sqrt{\dfrac{S_1^2}{n_1}+\dfrac{S_2^2}{n_2}}}<Z_{\alpha/2}\right]=1-\alpha$$

ومنه:

$$(\overline{X}_1-\overline{X}_2)-Z_{\alpha/2}\cdot\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}<(\mu_1-\mu_2)<(\overline{X}_1-\overline{X}_2)+Z_{\alpha/2}\cdot\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}$$

وهو مجال الثقة المطلوب حول $(\mu_1-\mu_2)$.

وللمقارنة بين $\mu_2-\mu_1$ نفترض أن شكل المجال السابق هــــو $a\leq\mu_1-\mu_2\leq b$ وبالتالي نميز ثلاث حالات .

-1 $\mu_1>\mu_2 \Leftarrow \mu_1-\mu_2>0 \Leftarrow a>0\ ,b>0$

-2 $\mu_1<\mu_2 \Leftarrow \mu_1-\mu_2<0 \Leftarrow a<0\ ,b<0$

-3 $\Leftarrow a<0\ ;\ b>0$ (باحتمال $(1-\alpha)$ يمكن أن يكــون $\mu_1-\mu_2=0$ أي $\mu_1=\mu_2$)

ثم نتخذ القرار على ضوء النتائج في المقارنة وطبيعة الدراسة.

مثال (5-6) :

أجريت دراسة في إحدى الجامعات للمقارنة بين المعدلات التي حصلت عليـــــها مجموعتان من الطلبة الأولى من المتزوجين والأخرى من غير المتزوجين، ولهذه الغايـــة أخذت عينتان واحدة من كل منها، وبعد إجراء الامتحان جاءت النتيجة كما يلي:

$S_1=4$	$\overline{X}_1=28.5$	$n_1=100$	σ_1^2	μ_1	مجتمـــــع المتزوجين
$S_2=3$	$\overline{X}_2=27.3$	$n_2=100$	σ_2^2	μ_2	مجتمع غـــير المتزوجين

عين 95% مجال ثقة الفرق الحقيقي بين متوسطي المجتمعين وماذا نستنتج؟

الحل:

إن $1-\alpha=0.95$ مجال ثقة حول $\mu_1-\mu_2$ يكون من الشكل :

$$(\bar{x}_1-\bar{x}_2)-Z_{\alpha/2}\cdot\sqrt{\frac{S_2^2}{n_1}+\frac{S_2^2}{n_2}}\le\mu_1-\mu_2\le(\bar{x}_1-\bar{x}_2)+Z_{\alpha/2}\cdot\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}$$

$$\bar{x}_1-\bar{x}_2=28.5-27.3=1.2$$

$$e=\left|\pm Z_{\alpha/2}\cdot\sqrt{\frac{S_1^2}{n_2}+\frac{S_2^2}{n_2}}\right|=(1.95)\sqrt{\frac{16}{100}+\frac{9}{100}}=0.98$$

ومنه :

$$1.2-0.98\le\mu_1-\mu_2\le1.2+0.98$$

$$0.22\le\mu_1-\mu_2\le2.18$$

وبما أن طرفي المجال موجبان عندئذ $\mu_1-\mu_2>0$ ومنه $\mu_1>\mu_2$ أي أنه بثقة 95% يكون متوسط الدرجات عند المتزوجين أفضل من متوسط الدرجات عند غير المتزوجين.

6-6: التقدير المجالي للفرق بين متوسطي مجتمعين يتوزعان على وجه التقريب طبيعيا وعينات صغيرة الحجم:

في هذه الحالة لدينا مجتمعان إحصائيان الأول توقعه μ_1 وتباين σ_1^2 الثاني توقعه μ_2 وتباينة σ_2^2 ونفترض أن:

$\sigma_1^2=\sigma_2^2=\sigma^2$ (خاصة التجانس) فعندئذ من أجل عينة من المجتمع الأول حجمها $30>n_1$ ومتوسطها $\bar{X}_2$ وتباينها S_1^2 وعينة من المجتمع الثاني مستقلة عن الأولى حجمها $30>n_2$ ومتوسطها $\bar{X}_2$ وتباينها S_1^2 فإن التباين المشترك S_P^2 والمعطى بالعلاقة:

$S_P^2=\dfrac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$ يستخدم في تقدير $\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}$ حيث المقدر يكون:

$S_P.\sqrt{\dfrac{1}{n_2}+\dfrac{1}{n_2}}$ وباستخدام توزيع ستودنت كون العينات صغيرة الحجم يكون الإحصاء:

$$T=\frac{(X_1-X_2)-(\mu_1-\mu_2)}{S_P\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim t_{(v=n_1+n_2-2)} \qquad \text{(ستودنت)}$$

وباستخدام العلاقة الاحتمالية في توزيع ستودنت:

$$P\left[-t_{\alpha/2}(v)<T<t_{\alpha/2}(v)\right]=1-\alpha$$

فإن $(1-\alpha)$ مجال ثقة حول $\mu_1-\mu_2$ يكون من الشكل (حسب (5-6):

$$(\overline{X}_1-\overline{X}_2)-Z_{\alpha/2}.S_p.\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\leq(\mu_1-\mu_2)\leq(\overline{X}_1-\overline{X}_2)+Z_{\alpha/2}.S_p.\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

أي أن $(1-\alpha)$ حدي ثقة حول $\mu_1-\mu_2$ هما

$$(\overline{X}_1-\overline{X}_2)\pm Z_{\alpha/2}.S_p.\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

ثم نجري المقارنة في ضوء المجال الناتج وطبيعة الدراسة (حسب (5.6))

مثال (6-6):

اختيرت مجموعتان من الطلاب $n_1=12$, $n_2=10$ وطبقت عليهما طريقتان مختلفتان في التعليم. وفي نهاية الفصل الدراسي أجري لهما اختبار واحد. وكان متوسط درجات المجموعة الأولى 85 بانحراف معياري قدره 4 بينما سجلت المجموعة الثانية متوسطا قدره 81 وبانحراف معياري قدره 5 عين 90% مجال ثقة حول الفرق الحقيقي بين متوسطي المجتمعين للمجموعتين وذلك بفرض أنهما موزعان على وجه التقريب طبيعيا ولهما التباين نفسه. وماذا نستنتج؟

الحل :

ليكن μ_1 متوسط مجتمع المجموعة الأولى و μ_2 متوسط مجتمع المجموعة الثانية ولنحسب:

$$S_P^2=\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}=\frac{(11)(16)+(9)(25)}{12+10-2}$$

$$\Rightarrow S_P^2=20.05\Rightarrow S_P=4.478$$

$$\overline{X}_1-\overline{X}_2=85-81=4$$

$$1-\alpha=0.90\Rightarrow\alpha=0.10\Rightarrow\frac{\alpha}{2}=0.05$$

$$\nu=n_1+n_2-2=12+10-2=20$$

$$t_{\alpha/2}(\nu)=t_{0.05}(20)=1.725$$ (من جدول ستودنت)

وبالتالي إن $1-\alpha$ مجال ثقة حول $(\mu_1-\mu_2)$:

$$(\overline{X}_1-\overline{X}_2)-t_{\alpha/2}.S_P.\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\leq\mu_1-\mu_2\leq(\overline{X}_1-\overline{X}_2)+t_{\alpha/2}.S_P.\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

$$4-(1.725)(4.478).\sqrt{\frac{1}{12}+\frac{1}{10}}\leq(\mu_1-\mu_2)\leq4+(1.725)(4.478).\sqrt{\frac{1}{12}+\frac{1}{10}}$$

$$0.69\leq\mu_1-\mu_2\leq7.31$$

وبما أن طرفي المجال موجبان عندئذ $\mu_1-\mu_2>0$ وبالتالي $\mu_1>\mu_2$ أي أن الطريقة الأولى أفضل من الطريقة الثانية وبثقة 90%.

7-6: التقدير المجالي للفرق بين وسيطي مجتمعين ثنائيين:

ليكن لدينا مجتمعان ثنائيان وسيطاهما P_1 , P_2 ولنفرض لدينا عينتين مستقلتين حجاهما n_1 , n_2 اخيرت الأولى من المجتمع الأول واختيرت الثانية من المجتمع الثاني بالمتوسطات n_1P_1 , n_2P_2 على الترتيب وبالانحراف المعياري $\sqrt{n_1P_1Q_1}$, $\sqrt{n_2P_2Q_2}$ على الترتيب. وسنرمز بـ $\hat{P}_1$, $\hat{P}_2$ لنسبتي النجاح في العينتين ويكون $\hat{P}_1-\hat{P}_2$ مقدر نقطي لـ P_1-P_2 حيث:

$$E(\hat{P}_1-\hat{P}_2)=E(\hat{P}_1)-E(\hat{P}_2)-P_1-P_2$$

$$V(\hat{P}_1-\hat{P}_2)=V(\hat{P}_1)+V(\hat{P}_2)=\frac{P_1Q_1}{n_1}+\frac{P_2Q_2}{n_2}$$

ومن أجل عينات كبيرة الحجم واستخدام تقريب التوزيع الطبيعي للتوزيع الحداني نجد أن:

$$(\hat{P}_1-\hat{P}_2)\sim N\left(P_1-P_2,\frac{P_1Q_1}{n_1}+\frac{P_2Q_2}{N_2}\right)$$

ومنه :

$$Z=\frac{(\hat{P}_1-\hat{P}_2)-(P_1-P_2)}{\sqrt{\frac{P_1Q_1}{n_1}+\frac{P_2Q_2}{n_n}}}\sim N(0,1)$$

وحسب (6-4):

$$\Leftrightarrow Z=\frac{(\hat{P}_1-\hat{P}_2)-(P_1-P_2)}{\sqrt{\frac{\hat{P}_1\hat{q}_1}{n_1}+\frac{\hat{P}_2\hat{q}_2}{n_n}}}\sim N(0,1)$$

وبالتالي من أجل $(1-\alpha)$ مجال ثقة حول (P_1-P_2) نستخدم العلاقة الاحتمالية في التوزيع الطبيعي المعياري.

$$P\left[-Z_{\alpha/2}\leq Z\leq Z_{\alpha/2}\right]=1-\alpha$$

$$\Rightarrow P\left[-Z_{\alpha/2}\leq\frac{(\hat{P}_1-\hat{P}_2)-(P_1-P_2)}{\sqrt{\frac{\hat{P}_1.\hat{q}_1}{n}+\frac{\hat{P}_2.\hat{q}_2}{n}}}\leq Z_{\alpha/2}\right]=1-\alpha$$

وبالتالي إن $(1-\alpha)$ مجال ثقة حول (P_1-P_2) يكون من الشكل:

$$(\hat{P}_1-\hat{P}_2)-Z_{\alpha/2}.\sqrt{\frac{\hat{P}_1\hat{q}_1}{n_1}+\frac{\hat{P}_2\hat{q}_2}{n_2}}\leq(P_1-P_2)\leq(\hat{P}_1-\hat{P}_2)+Z_{\alpha/2}.\sqrt{\frac{\hat{P}_1\hat{q}_1}{n_1}+\frac{\hat{P}_1\hat{q}_1}{n_2}}$$

ونجري المقارنة والاستنتاج على ضوء المجال الناتج تماما كما ورد في الفقرة (6-5).

مثال (6-7):

طبقت طريقتان A , B لمعالجة مرض معين فأخذت عينتان من المرضى، طبقت على الأولى الطريقة A وطبقت على الثانية الطريقة B، فإذا كان حجم العينة الأولى42= n_1 مريضا شفي منهم 18 وكان حجم العينة الثانية 38= n_2 مريضا شفي

منهم 15. والمطلوب تعيين 99% مجال ثقة حول الفرق الحقيقي بين نسبتي اللذين شفيوا في مجتمع العينتان وماذا نستنتج:

الحل:

لتكن P_1 النسبة الحقيقية للشفاء باستخدام الطريقة A

ولتكن P_2 النسبة الحقيقية للشفاء باستخدام الطريقة B.

ولدينا:

$$\hat{P}_1-\hat{P}_2=\frac{X_1}{n_1}-\frac{X_2}{n_2}=\frac{18}{42}-\frac{15}{38}=0.034$$

حيث : $\hat{P}_1=\frac{18}{42}=0.43 \quad ; \quad \hat{q}_1=0.57$

وحيث: $\hat{P}_2=\frac{15}{38}=0.39 \quad ; \quad \hat{q}_2=0.61$

ومن أجل $1-\alpha=0.99$ يكون $Z_{\alpha/2}=2.58$

ولدينا $(1-\alpha)$ مجال ثقة حول P_1-P_2 من الشكل:

$$(\hat{P}_1-\hat{P}_2)-Z_{\alpha/2}\cdot\sqrt{\frac{\hat{P}_1\hat{q}_1}{n_1}+\frac{\hat{P}_2\hat{q}_2}{n_2}}\le(P_1-P_2)\le(\hat{P}_1-\hat{P}_2)+Z_{\alpha/2}\cdot\sqrt{\frac{\hat{P}_1\hat{q}_1}{n_1}+\frac{\hat{P}_2\hat{q}_2}{n_2}}$$

ولدينا:

$$Z_{\alpha/2}\cdot\sqrt{\frac{\hat{P}_1\hat{q}_1}{n_1}+\frac{\hat{P}_2\hat{q}_2}{n_2}}=(2.58)\cdot\sqrt{\frac{(0.43)(0.57)}{42}+\frac{(0.39)(0.61)}{38}}$$
$$=0.28$$

وبالتالي $(1-\alpha)=0.99$ مجال ثقة حول P_1-P_2

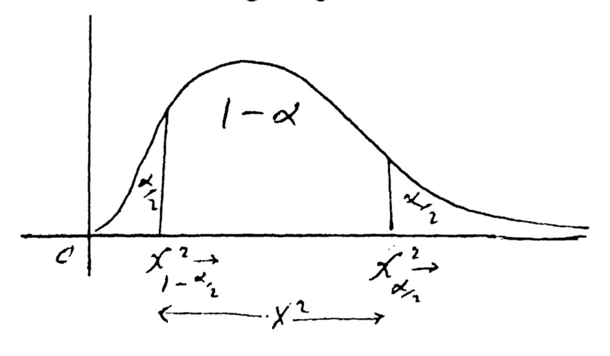
$$0.034-0.28\le(P_1-P_2)\le0.034+0.28$$
$$-0.246\le(P_1-P_2)\le0.314$$

وبما أن طرفي المجال الناتج الأدنى سالب والأعلى موجب فهذا يعني أنه بثقة 99% يكون $P_1 - P_2 = 0$ أي $P_1 = P_2$ وبالتالي لا فرق بين طريقتي العلاج للشفاء من المرض.

8-6: التقدير المجالي للتباين:

ليكن لدينا مجتمع إحصائي يتوزع على وجه التقريب طبيعيا بتوقع μ وتباين σ^2 ولنسحب من هذا المجتمع عينة عشوائية من الحجم N. رأينا في (مقرر مبادىء الإحصاء النظري) أن الإحصاء $X^2 = \frac{(n-1)S^2}{\sigma^2}$ يتوزع وفق توزيع كاي مربع بدرجة من الحرية $v = n-1$. (حيث S^2 تباين العينة المسحوبة).

ورأينا أن العلاقة الاحتمالية في توزيع كاي مربع تكون من الشكل:



$$P\left[X^2_{1-\alpha/2}(v) \le X^2 \le X^2_{\alpha/2}(v)\right] = 1-\alpha$$

ومن أجل بناء $(1-\alpha)$ مجال ثقة حول σ^2 نبدل X^2 بما يساويها في الاحتمال الأخير:

$$P\left[X^2_{1-\alpha/2}(\nu)\leq\frac{(n-1)S^2}{\sigma^2}\leq X^2_{\alpha/2}(\nu)\right]=1-\alpha$$

$$\Rightarrow P\left[\frac{(n-1)S^2}{X^2_{\alpha/2}(\nu)}\leq\sigma^2\leq\frac{(n-1)S^2}{X^2_{1-\alpha/2}(\nu)}\right]=1-\alpha$$

ومنه $(1-\alpha)$ مجال ثقة حول σ^2 يكون من الشكل:

$$\frac{(n-1)S^2}{X^2_{\alpha/2}(\nu)}\leq\sigma^2\leq\frac{(n-1)S^2}{X^2_{1-\alpha/2}(\nu)}$$

ومنه نستنتج $(1-\alpha)$ مجال ثقة حول الانحراف المعياري σ:

$$\sqrt{\frac{(n-1)S^2}{X^2_{\alpha/2}(\nu)}}\leq\sigma\leq\sqrt{\frac{(n-1)S^2}{X^2_{1-\alpha/2}(\nu)}}$$

حيث $X^2_{1-\alpha/2}(\nu)$ و $X^2_{\alpha/2}(\nu)$ تستنتج من جدول توزيع كاي مربع والذي يحسب المساحات من اليمين.

مثال (8-6):

لدينا عشر علب للأغذية المحفوظة أوزانها بالأونزات:

16.0; 16.4; 16.1; 15.8; 17.0; 16.1; 15.9; 15.9; 16.9; 15.2

1- عين 95% مجال ثقة حول التباين الحقيقي σ^2 للأوزان.

2- عين 95% مجال ثقة حول الانحراف المعياري الحقيقي σ للأوزان.

الحل:

1- نحسب أولا تباين العينة:

$$S^2=\frac{n\sum_{i=1}^{n}X_i^2-\left(\sum_{i=1}^{n}X_i\right)^2}{n(n-1)}$$

$$=\frac{10(260112)-(161.2)^2}{10(10-1)}=0.286$$

ومن أجل

$$1-\alpha=0.95 \Rightarrow \alpha=0.05 \Rightarrow \frac{\alpha}{2}=0.025$$

$$\nu=n-1=10-1=9$$

$$X^2_{\alpha/2}(\nu)=X^2_{0.025}(9)=19.0228$$

$$X^2_{1-\alpha/2}(\nu)=X^2_{0.975}(9)=2.70039$$

ومنه $(1-\alpha=0.95)$ مجال ثقة حول σ^2:

$$\frac{(n-1)S^2}{X^2_{\alpha/2}(\nu)} < \sigma^2 < \frac{(n-1)S^2}{X^2_{1-\alpha/2}(\nu)}$$

$$\frac{(9)(0.286)}{19.0228} < \sigma^2 < \frac{(9)(0.286)}{2.70039}$$

$$0.135 < \sigma^2 < 0.953$$

وبالتالي بثقة 95% فإن تباين الوزن لن يقل عن 0.135 ولن يزيد على 0.953.

2- وإن $(1-\alpha=0.95)$ حول الانحراف المعياري للوزن يكون من الشكل:

$$\sqrt{0.135} < \sigma < \sqrt{0.953}$$

$$0.367 < \sigma < 0.976$$

أي نكون واثقين 95% من أن الانحراف المعياري للوزن لن يقل عن 0.367 ولن يزيد على 0.976.

9-6: التقدير المجالي للنسبة بين تباينين:

قد نرغب أحيانا في مقارنة دقة جهاز قياس بدقة جهاز قياس آخر أو استقراء كل من خطين للإنتاج في صناعة معينة وأمثله عديدة في ذلك المجال فإن ذلك يمكن إجراؤه في دراسة المقارنة بين تبايني مجتمعين بوساطة النسبة بينهما.

فإذا كان لدينا مجتمعان طبيعيان توقع الأول μ_1 وتباينة σ_1^2 وتوقع الثاني μ_2 وتباينه σ_2^2 ثم نسحب من المجتمع الأول عينة عشوائية حجمها n_1 وتباينها S_1^2 ونسحب من المجتمع الثاني عينة عشوائية مستقلة عن الأولى حجمها n_2 وتباينها S_2^2 عندئذ يكون:

الإحصاء: $\frac{(n_1-1)S_1^2}{\sigma_1^2}$ يتوزع وفق كاي مربـــع بدرجـــة مـــن الحريـــة $V_1=(n_1-1)$ والإحصاء $\frac{(n_2-1)S_2^2}{\sigma_2^2}$ يتوزع وفق كاي مربع بدرجة مـــن الحريـــة $V_2=n_2-1$. وحسب تعريف توزيع فيشر من أجل مجتمعـــين مســـتقليين يكـــون الإحصاء:

(فيشر)

$$F=\frac{\frac{(n_1-1)S_1^2}{\sigma_1^2}\Big/(n_1-1)}{\frac{(n_2-1)S_2^2}{\sigma_2^2}\Big/(n_2-1)}\sim f_{(\nu_1=n_1-1,\,\nu_2=n_2-1)}$$

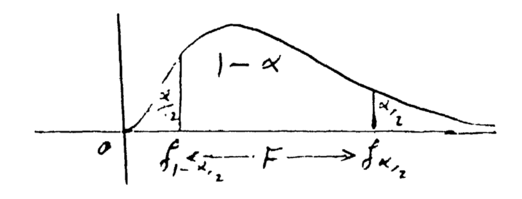
حيث ν_1 درجة حرية البسط و ν_2 درجة حرية المقام

أي :

$$F=\frac{S_1^2}{S_2^2}.\frac{\sigma_2^2}{\sigma_1^2}\sim f_{(\nu_1,\nu_2)}$$ (فيشر)

وحسب العلاقة الاحتمالية في توزيع فيشر:

$$P\left[f_{1-\alpha/2}(\nu_1,\nu_2)<F<f_{\alpha/2}(\nu_1,\nu_2)\right]=1-\alpha$$



ونستبدل بـــ T ما يساويها في العلاقة الاحتمالية نجد :

$$P\left[f_{1-\alpha/2}(\nu_1,\nu_2) < \frac{S_1^2}{S_2^2}\cdot\frac{\sigma_2^2}{\sigma_1^2} < f_{\alpha/2}(\nu_1,\nu_2)\right] = 1-\alpha$$

ومنه نجد:

$$P\left[\frac{S_1^2}{S_2^2}\frac{1}{f_{1-\alpha/2}(\nu_1,\nu_2)} < \frac{\sigma_2^2}{\sigma_1^2} < \frac{S_1^2}{S_2^2} < f_{\alpha/2}(\nu_1,\nu_2)\right] = 1-\alpha$$

أي أن $(1-\alpha)$ مجال ثقة حول $\frac{\sigma_1^2}{\sigma_2^2}$ يكون من الشكل:

$$\frac{S_1^2}{S_2^2}\frac{1}{f_{1-\alpha/2}(\nu_1,\nu_2)} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2} < \frac{1}{f_{1-\alpha/2}(\nu_1,\nu_2)}$$

وباستخدام الخاصة الأساسية في حساب الاحتمالات لتوزيع فيشر:

$$\frac{1}{f_{1-\alpha/2}(\nu_1,\nu_2)} = f_{\alpha/2}(\nu_1,\nu_2)$$

يصبح $(1-\alpha)$ مجال ثقة حول $\frac{\sigma_1^2}{\sigma_2^2}$ يكون من الشكل:

$$\frac{S_1^2}{S_2^2}\frac{1}{f_{1-\alpha/2}(\nu_1,\nu_2)} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2}\cdot f_{\alpha/2}(\nu_1,\nu_2)$$

وللمقارنة نميز ثلاث حالات:

1- $\sigma_1^2 > \sigma_2^2 \Leftarrow \frac{\sigma_1^2}{\sigma_2^2} > 1$

2- $\sigma_1^2 < \sigma_2^2 \Leftarrow \frac{\sigma_1^2}{\sigma_2^2} < 1$

3- إذا كانت $\frac{\sigma_1^2}{\sigma_2^2}$ بين قيمتين أكبر من الواحد وأقل من الواحد. فعندئذ من الممكن أن تكون $\sigma_1^2 = \sigma_2^2$ باحتمال $(1-\alpha)$ ثم نتخذ القرار على ضوء نتيجة النسبة وطبيعة الدراسة.

مثال (6-9):

في دراسة أجريت للمقارنة بين كمية النيكوتين في صنفين من الدخان A, B فإن السجائر من النوع A تحوي نيكوتينا بمعدل 3.1 وبانحراف معياري m.G 0.5. ومن أجل السجائر من النوع B فتبين أنها تحوي نيكوتينا بمعدل 2.7 وبانحراف معياري m.G 0.7 وذلك من أجل عينة من 10 سجائر من النوع A و8 سجائر من النوع B. وبفرض أن مجتمعي العينتين يتوزعان طبيعيا بتباين مختلف عين 98% مجال ثقة للنسبة الحقيقية للتباين في مجتمعين العينتين وماذا نستنتج من هذه الدراسة.

الحل :

إن $1-\alpha=0.98$ مجال ثقة حول $\frac{\sigma_1^2}{\sigma_2^2}$ حيث σ_1^2 هو التباين الحقيقي لكمية النيكوتين في النوع A و σ_2^2 هو التباين الحقيقي لكمية النيكوتين في النوع B من الشكل:

$$\frac{S_1^2}{S_2^2}\cdot\frac{1}{f_{\alpha/2}(n_1-1,n_2-2)}<\frac{\sigma_1^2}{\sigma_2^2}<\frac{S_1^2}{S_2^2}f_{\alpha/2}(n_2-1,n_1-1)$$

حيث :

$$1-\alpha=0.98\Rightarrow\alpha=0.02\ ;\ \tfrac{\alpha}{2}=0.01$$

$$v_1=n_1-1=10-1=9;\ v_2=n_2-1=8-1=7$$

(من جدول فيشر): $f_{\alpha/2}(n_1-1,n_2-1)=f_{0.01}(9,7)=6.72$

$$f_{\alpha/2}(n_2-1,n_1-1)=f_{0.01}(7,9)=5.61$$

ولدينا:

$$S_1^2=(0.5)^2=0.25\quad;\ S_2^2=(0.7)^2=0.49$$

ومنه $1-\alpha=0.98$ مجال ثقة حول $\frac{\sigma_1^2}{\sigma_2^2}$:

$$\left(\frac{0.25}{0.49}\right)\cdot\frac{1}{6.72}<\frac{\sigma_1^2}{\sigma_2^2}<\frac{0.25}{0.49}(5.61)$$

$$0.076<\frac{\sigma_1^2}{\sigma_2^2}<2.862$$

وبما أن النسبة محصورة بين قيمتين تحوي القيمة واحد فإنه من الممكن أن يكـــون $\frac{\sigma_1^2}{\sigma_2^2}=1$ أي $\sigma_1^2=\sigma_2^2$ أي لا فرق حقيقي بين التباينين.

10-6: اختبار الفرضيات:

تعرف الفرضية الإحصائية بأنها مقولة أو إفادة تتعلق بالمجتمع الإحصائي وتحتمـــل الصحة أو الخطأ، وإن صحة الفرضية أو خطأها لا يمكن معرفتها بدقة إلا إذا تفحصنا المجتمع بأكمله وهذا أمر غير عملي. ولذلك نختار عينة عشوائية من هـــذا المجتمـــع، ونستخدم المعلومات التي تحويها العينة لنقرر ما إذا كانت الفرضية الإحصائية صحيحة أم خاطئة والخطوات الرئيسة لاختبار الفرضية تكون كما يلي:

1- نسمي الفرضية الابتدائية حول وسيط مجتمع إحصائي θ ونقول $H_0:\theta=\theta_0$

ونسمي الفرضية البديلة H_1 والتي سنقلبها في حال رفض H_0 حيث تكون H_1:

إما $H_1:\theta>\theta_0$ ويدعى باختبار من الطرف الأيمن

أو $H_1:\theta<\theta_0$ ويدعى باختبار من الطرف الأيسر

أو $H_1:\theta\neq\theta_0$ ويدعى باختبار من الطرفين

2- تتخذ قاعدة لاتخاذ القرار ، تقبل في ضوئها الفرضية أو نرفضها، لذلك نعـرف خطأين:

خطأ من النوع الأول: وهو الخطأ الناجم من رفض الفرضية الابتدائية H_0 علما بأنها صحيحة ونرمز بـــ α لاحتمال رفض H_0 وهي صحيحـــة. ونســـمي α مستوى أهمية الاختبار. حيث نختبر H_0 مقابل H_1.بمستوى الدلالة α.

خطأ من النوع الثاني: وهو الخطأ الناجم من قبول الفرضية الابتدائية H_0 وهـــي خاطئة ونرمز لاحتمال القبول الخاطيء بـــ β.

3- نختار عينة عشوائية من المجتمع الإحصائي المدروس ونحسب إحصائيات هـــذه العينة علما بأن H_0 صحيحة مع الأخذ بالحسبان توزيع العينة (توزيـــع طبيعـــي – ستودنت – كاي مربع – فيشر...) وندعو الإحصائية المحسوبة تحت صحـــة H_0 بإحصاء الاختبار.

4- اعتمادا على مستوى الأهمية α ونوع توزيع العينة ونوع الاختبار (طــــرف أيمن، طرف أيسر، طرفين).

نحدد مناطق الرفض والقبول لــ α حيث ننشيء منطقة رفض مســـاحتها H_0 تكون:

من جهة اليمين إذا كان الاختبار من الطرف الأيمن

ومن جهة اليسار إذا كان الاختبار من الطرف الأيسر

ومن الطرفيين إذا كان الاختبار من الطرفيين حيث مساحة كل طرف $\frac{\alpha}{2}$ بسبب تناظر التوزيع.

5- نقارن قيمة إحصاء الاختبار المحسوبة من (3) مع منـــاطق الرفــض والقبـــول والمحددة من (4). فإذا وقع إحصاء الاختبار في منطقة القبول، فقبــــل عندئـــذ H_0 ونرفض H_1 وإذا وقع إحصاء الاختبار في منطقة الرفض، فنرفض H_0 ونقبل H_1.

وأغلب الاختبارات التي تجرى، تجري من الطرفين ونتيجة المقارنة في (5) يتحـــدد لنا جهة الرفض وبالتالي فالاختبار من الطرفين يحوي الاختبار من طرف واحد.

11-6: اختبارات حول μ (توقع مجتمع إحصائي):

ليكن لدينا مجتمع إحصائي توقعه μ وتباينه σ^2 (معلوم أو يمكن تقديره مـــن تباين عينة مسحوبة منه S^2 من حجم معين n) .

والمطلوب: اختبار الفرضية $H_0: \mu = \mu_0$ مقابل الفرضية

البديلة $H_1: \mu > \mu_0$ اختبار من اليمين.

أو $H_1: \mu < \mu_0$ اختبار من اليسار

أو $H_1: \mu \neq \mu_0$ اختبار من الطرفين

عند مستوى الدلالة α (احتمال رفض H_0 وهي صحيحة).

ثم نسحب عينة عشوائية من الحجم n وإحصاء الاختبار يكون

$$Z = \frac{X - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

سواء أكان المجتمع طبيعيا وعينة كبيرة الحجم أو بتطبيق مبرهنه النهاية المركزيـــة .(2.6)

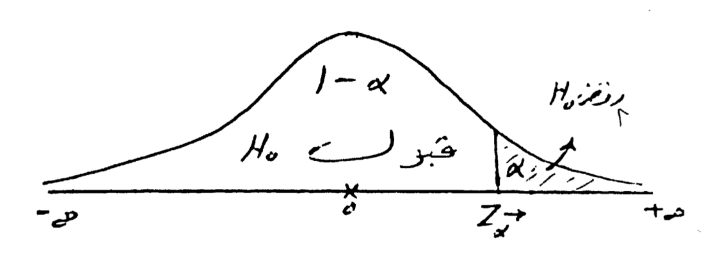
ثم نحسب قيمة إحصاء الاختبار عندما تكون H_0 صحية أي

$$Z_0 = \frac{\overline{X} - \mu_0}{\frac{s}{\sqrt{n}}} \; ; \; S = \sqrt{\frac{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}{n(n-1)}}$$

ثم نحدد مناطق الرفض والقبول حسب قيمة α ونوع التوزيع ونوع الاختبار.

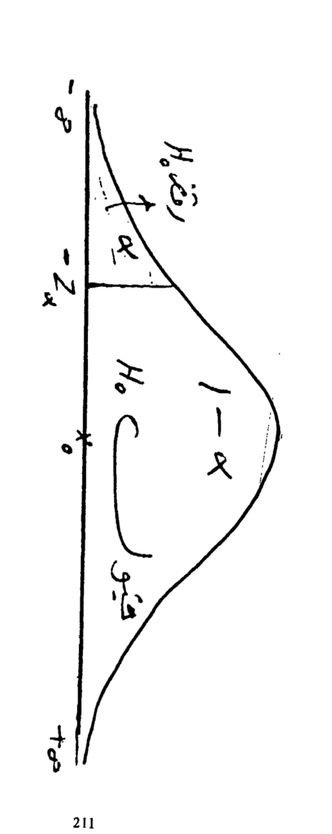
فمن أجل اختبار من اليمين : أي $H_0 : \mu = \mu_0$ مقابل $H_0 : \mu > \mu_0$ فتكون منطقة القبول $]-\infty, Z_x[$ ومنطقة الرفض $]Z_x, +\infty[$

ومن أجل اختبار من اليسار أي :



$H_0 : \mu = \mu_0$ مقابل $H_0 : \mu < \mu_0$ تكون منطقة القبول $]-Z_\alpha, +\infty[$

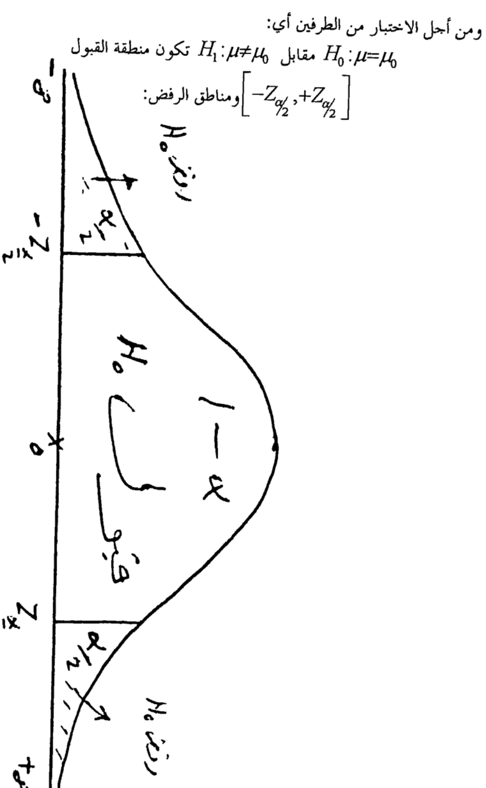
ومنطقة الرفض $]-\infty, -Z_\alpha[$



-∞
α
-Z α
H0
0
1 - α
+∞

ومن أجل الاختبار من الطرفين أي:

$H_0 : \mu = \mu_0$ مقابل $H_1 : \mu \neq \mu_0$ تكون منطقة القبول $\left[-Z_{\alpha/2}, +Z_{\alpha/2}\right]$ ومناطق الرفض:



$$\left]-\infty, -Z_{\alpha/2}\right[\quad و \quad \left]Z_{\alpha/2}, +\infty\right[$$

حيث تتحدد Z_α و $Z_{\alpha/2}$ من جدول التوزيع الطبيعي المعياري كما رأينا سابقا.

ثم نقارن Z_0 مع مناطق الرفض والقبول فإذا وقعت Z_0 في منطقة القبول فهذا يعني أننا ستقبل H_0 ونرفض H_1. وإذا وقعت Z_0 في منطقة الرفض فهذا يعني أننا نقبل H_1 ونرفض H_0.

مثال (6-10):

تبين من عينة عشوائية حجمها n=100 متوفى أن متوسط العمر لهؤلاء هو 71.8 سنة بانحراف معياري قدره 8.9 سنة. فهل يشير هذا إلى أن مستوى العمر الآن أكبر من 70 سنة بمستوى الأهمية $\alpha=0.05$؟

الحل:

الفرضية الابتدائية: $H_0: \mu=70$

والفرضية البديلة: $H_1: \mu>70$

ومستوى الدلالة الإحصائية (الأهمية) : $\alpha=0.05$

وإحصاء الاختبار $Z=\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$

ثم نحسب قيمة إحصاء الاختبار عندما تكون H_0 صحيحة

$$Z=\frac{X-\mu_0}{\frac{S}{\sqrt{n}}}=\frac{71.8-70}{\frac{8.9}{\sqrt{100}}}=2.02$$

وبناء على $\alpha=0.05$ والتوزيع طبيعي معياري والاختبار من الطرف الأيمن نحدد منطقة الرفض والقبول حيث من أجل $\alpha=0.05$ فإن $Z_\alpha=1.645$ وبالتالي:

$\left]-\infty, Z_\alpha\right] = \left]-\infty, 1.645\right]$: (منطقة القبول)

$]Z_\alpha,+\infty[=]1.645,+\infty[$: (منطقة الرفض)

وبمقارنة $Z_0=0.02$ مع مناطق الرفض والقبول نجد أن $Z_0\in]1.645,+\infty[$ أي تقع في منطقة الرفض. ومنه نرفض H_0 ويقتل H_1، أي أن مستوى العمر الآن أكثر فعلا من 70.

مثال (6-11):

لنفترض أن المعدل الوسطي لدرجات مجتمع من الطلبة هو 110 بانحراف معياري 10. أخذنا منه عينة من 49 طالبا من الذكور فقط. وحسبنا المعدل الوسطي لدرجاتهم، فكانت 112 درجة. فهل تدل هذه النتيجة على اختلاف معدلات الذكور بمستوى 0.05 من الأهمية.

الحل:

الفرضية الابتدائية: $H_0:\mu=100$

الفرضية البديلة: $H_1:\mu\neq110$

ومستوى الدلالة: $\alpha=0.05$ والاختبار من الطرفين وإحصاء الاختبار

$$Z=\frac{X-\mu}{\frac{\sigma}{\sqrt{n}}}\sim N(0,1)$$

وعندما تكون H_0 صحيحة، نحسب قيمة إحصاء الاختبار:

$$Z_0=\frac{X-\mu_0}{\frac{S}{\sqrt{n}}}=\frac{112-110}{\frac{10}{\sqrt{49}}}=1.40$$

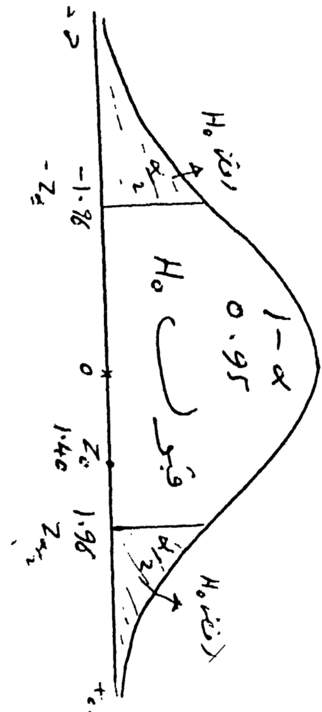
وحسب $\alpha=0.05$ والاختبار من الطرفين والتوزيع طبيعي ميعاري تكون منطقة القبول.

$$\left[-Z_{\alpha/2},Z_{\alpha/2}\right]=[-1.96,1.96]$$

ومنطقة الرفض:

$$\left]-\infty-Z_{\alpha/2}\right[=]-\infty,-1.96[$$

$$\left]Z_{\alpha/2},+\infty\right[=]1.96,+\infty[$$



وبمقارنة $Z_0=1.40$ مع مناطق الرفض والقبول نجد أن Z_0 تقع في منطقة القبول (انظر الشكل) وبالتالي نقبل H_0 ونرفض H_1 أي بثقة 95% وبخطأ 0.05 لا يوجد اختلاف في معدلات الذكور.

12-6: الاختبارات الخاصة بوسيط مجتمع ثنائي:

إن اختبار الفرضيات المختصة بالنسبة P مرغوبة لدى العديد من مجالات الحياة. فرجال السياسة يهتمون بمعرفة نسبة المقترعين لصالحهم، والتاجر يهتم بمعرفة نسبة التلفيات في البضاعة المشحونة، وصاحب المصنع يهتم بمعرفة نسبة القطع المعيبة الصنع في إنتاجه وهكذا. وهنا نختبر الفرضية الابتدائية: $H_0: P=P_0$ مقابل الفرضية البديلة $H_1: P>P_0$ أو $H_1: P<P_0$ أو $H_1: P\neq P_0$ وعند مستوى الدلالة α.

إن إحصاء الاختبار المستخدم هو:

$$Z=\frac{\hat{P}-P}{\sqrt{\frac{PQ}{n}}}\sim N(0,1)$$

(حسب ما رأينا في (4.6)).

حيث $\hat{P}=\frac{X}{n}$: (X تمثل عدد النجاحات الحاصلة من n تكرارا مستقلا)

ثم نحسب قيمة إحصاء الاختبار عندما تكون H_0 صحيحة أي:

$$Z_0=\frac{\hat{P}-P_0}{\sqrt{\frac{P_0Q_0}{n}}}=\frac{X-nP_0}{\sqrt{nP_0Q_0}}$$

ثم نتابع الاختبار تماما كما ورد في (11.6).

حيث نحدد مناطق الرفض والقبول ونجري المقارنة بينها وبين Z_0 ثم نتخذ القرار المناسب.

مثال (12-6) :

من المعلوم أن 54% من سكان مدينة معينة يفضلون السكن في الطوابق العليا، ونظرا لظروف الصيانة السيئة للمصاعد والأعطال المتكررة فيها، يعتقد أن تغييرا قد طرأ على هذه النسبة. ولاختبار هذا التغير ، أخذت عينة من 1000 شخص، فكان 480 منهم يفضلون السكن في الطوابق العليا ، فهل تدعم هذه النتيجة هذا الاعتقاد وذلك بمستوى الدلالة $\alpha=0.01$.

الحل:

الفرضية الابتدائية $H_0: P=0.54$

الفرضية البديلة $H_1: P \neq 0.54$

ومستوى الأهمية $\alpha = 0.01$ والاختبار من الطرفين وإحصاء الاختبار

$$Z = \frac{\hat{P} - P}{\sqrt{nPQ}} \sim N(0,1)$$

ثم نحسب قيمة إحصاء الاختبار عندما تكون H_0 صحيحة أي:

$$Z_0 = \frac{X - nP_0}{\sqrt{nP_0 Q_0}} = \frac{480 - (1000)(0.54)}{\sqrt{(1000)(0.54)(0.46)}}$$

$$= -3.82$$

واعتمادا على مستوى الدلالة $\alpha = 0.01$ والاختبار من الطرفين والتوزيع طبيعي معياري تكون منطقة القبول.

$$\left[-Z_{\alpha/2}, Z_{\alpha/2} \right] = [-2.58, 2.58]$$

ومنطقة الرفض:

$$\begin{cases} \left] -\infty, -Z_{\alpha/2} \right[=]-\infty, -2.58[\\ \left] Z_{\alpha/2}, +\infty \right[=]2.58, +[\end{cases}$$

ثم نقارن $Z_0 = -3.82$ مع مناطق الرفض والقبول فنجد أن:

$$Z_0 \in]-\infty, -2,58[$$

أي تنتمي إلى منطقة الرفض من جهة اليسار وبالتالي نرفض H_0 وتقبل H_1 أي أن النسبة الحقيقية قد تغيرت واتجهت نحو الأقل (لأن الرفض من جهة اليسـلر) أي أن الذين يفضلون السكن في الطوابق العليا الآن هم أقل مما قبل.

13-6: الاختبارات الخاصة بالفرق بين متوسطي مجتمعين إحصـــــائيين: $(\mu_1 - \mu_2)$:

قد نرغب أحيانا في اختبار فرضية حول متوسطي مجتمعين مختلفين أي اختيــــار الفرضية الابتدائية $H_0: \mu_1 - \mu_2 = d$ (حيث d ثابت وعلى الغـــالب d=0). مقـــابل

الفرضية البديلة $H_1: \mu_1 - \mu_2 > d$ أو $H_1: \mu_1 - \mu_2 < d$ أو $H_1: \mu_1 - \mu_2 = d$ وعند مستوى الأهمية α.

يكون إحصاء الاختبار :

$$Z = \frac{(X_1 - X_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

(تماما كما ورد في (15.6))

ثم نحسب قيمة إحصاء الاختبار عندما تكون H_0 صحيحة

$$Z_0 = \frac{(X_1 - X_2) - d}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

ثم اعتمادا على α والتوزيع طبيعي معياري ونوع الاختبار نحدد مناطق الرفض والقبول ونجري المقارنة مع Z_0 ونتخذ القرار المناسب (كما ورد في (6-11)).

مثال (6-13):

في اختبار تجريبي في مادة الإحصاء ، تقدم 75 طالبا و50 طالبة. فكان متوسط درجات الطلاب 82 بانحراف معياري 8 درجات، بينما كان متوسط درجات الطالبات 76 بانحراف معياري 6 درجات.

والمطلوب اختبار فيما إذا كان الطلاب والطالبات يعملون بالسوية ذاتها في الاختبار التجريبي وذلك عند مستوى الدلالة الإحصائية α=0.05 .

الحل:

الفرضية الابتدائية $H_0: \mu_1 - \mu_2 = 0$

الفرضية البديلة $H_0: \mu_1 - \mu_2 \neq 0$

حيث μ_1 متوسط درجات مجتمع الطلاب

و μ_2 متوسط درجات مجتمع الطالبات

وعند مستوى الدلالة $\alpha = 0.05$ والاختبار من الطرفين يكون إحصاء الاختبار

$$Z=\frac{(X_1-X_2)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}\sim N(0,1)$$

وتكون قيمته عندما تكون H_0 صحيحة

$$Z=\frac{(X_1-X_2)-0}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}}=\frac{(82-76)}{\sqrt{\frac{64}{75}+\frac{36}{50}}}=4.78$$

واعتمادا على مستوى الدلالة $\alpha=0.05$ والتوزيع طبيعي معياري والاختبار من الطرفين تكون منطقة القبول

$$\left[-Z_{\alpha/2},Z_{\alpha/2}\right]=[-1.96,1.96]$$

وتكون منطقة الرفض:

$$\begin{cases}\left]-\infty,-Z_{\alpha/2}\right[=\left]-\infty,-1.96\right[\\ \left]Z_{\alpha/2},+\infty\right[=\left]1.96,+\infty\right[\end{cases}$$

وبمقارنة $Z_0=4.78$ مع مناطق الرفض والقبول نجد أن

$$Z_0\in\left]1.96\,,+\infty\right[$$

أي أن Z_0 تقع في منطقة الرفض من جهة اليمين وبالتالي نرفض H_0 ونقبل H_1 أي أن هناك اختلافا في مستوى درجات الطلاب من مستوى درجات الطالبات وكون الرفض من اليمين فهذا يعني أن $\mu_1-\mu_2>0$ أي $\mu_1>\mu_2$ وبالتالي مستوى درجات الطلاب أفضل من مستوى درجات الطالبات. وهذه النتيجة تكون بثقة 95% وبخطأ 0.05.

14-6: اختبارات حول الفرق بين نسبتي مجتمعين ثنائيين:

تبرز هذه القضية إذا رغبنا في اختبار فرضية تساوي نسبتين، فمثلا يمكن لمدخن أن يقرر ما إذا كان عليه أن يقلع عن التدخين إذا اقتنع أن نسبة المدخنين المصابين بورم رئوي تفوق نسبة غير المدخنين والمصابين بهذا المرض. وفي الحالة العامة من أجل مجتمعين ثنائيين وسطاؤهما P_2,P_1 على الترتيب. نرغب في اختبار الفرضية الابتدائية

$H_0 : P_1 = P_2$ مقابل الفرضية البديلة: $H_1 : P_1 > P_2$ أو $H_1 : P_1 < P_2$ أو $H_1 : P_1 \neq P_2$.
وعند مستوى الدلالة α.
يكون إحصاء الاختبار

$$Z = \frac{(\hat{P}_1 - \hat{P}_2) - (P_1 - P_2)}{\sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}}} \sim N(0,1)$$

تماما كما رأينا في (7.6).
ونحسب قيمة إحصاء الاختبار عندما تكون H_0 صحيحة أي

$P_1 = P_2 = P$ أي $P = \frac{X_1 + X_2}{n_1 + n_2}$

حيث X_1 يمثل عدد النجاحات الحاصلة من n تكرارا من المجتمع P_1. و X_2 يمثل عدد النجاحات الحاصلة من n_2 تكرارا في المجتمع P_2 ومنه.

$$Z_0 = \frac{\hat{P}_1 - \hat{P}_2}{\sqrt{P.Q\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

واعتمادا على مستوى الدلالة ونوع الاختبار والتوزيع طبيعي معياري، نحدد مناطق الرفض والقبول ونقارن مع Z_0 ونتخذ القرار المناسب (كما رأينا في (11.6)).

مثال: (14-6):

يدعى باحث أن إقبال أبناء الأسر الغنية على الأعمال الحرة أكبر منه في الأسر الفقيرة ولاختبار هذا الادعاء أخذت عينتان من الشبان حجم كل منهما 1000، الأولى من أبناء الأسر ذوي الدخل العالي والأخرى من أبناء الأسر ذوي الدخل الصغير والمحدود.

فتبين أن نسبة المتوجهين إلى الأعمال الحرة في العينة الأولى هي 0.253 وفي الثانية هي 0.22.

فهل هناك ثمة فرق حقيقي بين النسبتين بمستوى 0.05 من الأهمية الإحصائية.

الحل:

لتكن P_1 النسبة الحقيقية لمن يفضلون الأعمال الحرة عند أبناء الأسر الغنية.

و P_2 النسبة الحقيقية لمن تفضلون الأعمال الحرة عند أبناء الأسر الفقيرة.

الفرضية الابتدائية: $H_0 : P_1 = P_2$

والفرضية البديلة : $H_0 : P_1 \neq P_2$

ومستوى الدلالة $\alpha_1 = 0.05$ والاختبار من الطرفين. وإحصاء الاختبار هو:

$$Z = \frac{(\hat{P}_1 - \hat{P}_2) - (P_1 - P_2)}{\sqrt{\frac{P_1Q_1}{n_1} + \frac{P_2Q_2}{n_2}}} \sim N(0,1)$$

ثم نحسب قيمة إحصاء الاختبار عندما تكون H_0 صحيحة:

$$Z_0 = \frac{\hat{P}_1 - \hat{P}_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$P = \frac{X_1 + X_2}{n_1 + n_2}; X_1 = n_1\hat{P}_1 = (1000)(0.253) = 253$$

$$X_2 = n_2\hat{P}_2 = (1000)(0.220) = 220$$

$$\Rightarrow P = \frac{253+220}{1000+1000} = \frac{473}{2000} = 0.237$$

$$Q = 1 - P = 1 - 0.237 = 0.763$$

$$Z_0 = \frac{0.253 - 0.22}{\sqrt{(0.237)(0.763)\left(\frac{1}{1000} + \frac{1}{1000}\right)}}$$

$$= \frac{0.033}{0.019} = 1.737$$

واعتمادا على مستوى الدلالة $\alpha = 0.05$ والاختبار من الطرفين والتوزيع طبيعـــي معياري تكون منطقة القبول

$$\left[-Z_{\alpha/2}, Z_{\alpha/2}\right]=[-1.96,1.96]$$

ومنطقة الرفض:

$$\begin{cases} \left]-\infty,-Z_{\alpha/2}\right[=\left]-\infty,-1.96\right[\\ \left]Z_{\alpha/2},+\infty\right[=\left]1.96,+\infty\right[\end{cases}$$

وبمقارنة $Z_0=1.737$ مع مناطق الرفض والقبول نجد أن $Z_0\in[-1.96,1.96]$ أي في منطقة القبول وبالتالي نقبل H_0 أي لا فرق بين نسبة من يفضلون الأعمال الحرة من أبناء الأسر الفنية ونسبة من يفضلون الأعمال الحرة من أبناء الأسر الفقيرة.

15-6: اختيار فرضيات حول متوسط مجتمع إحصائي يتوزع على وجه التقريب طبيعيا وعينات صغيرة الحجم:

بالعودة إلى الفقرة (6-3) من أجل عينة صغيرة الحجم أي n< 30 ، ومن أجل اختبارات تتعلق بمتوسط مجتمع إحصائي μ عندما يكون σ^2 مجهولا.

نريد أن نختبر الفرضية الابتدائية $H_0:\mu=\mu_0$ مقابل الفرضية البديلة: $H_1:\mu>\mu_0$ أو $H_1:\mu<\mu_0$ أو $H_1:\mu\neq\mu_0$ وعند مستوى الدلالة α

يكون إحصاء الاختبار:

$$T=\frac{X-\mu}{\frac{S}{\sqrt{n}}}\sim t_{(v=n-1)}^{(\text{ستودنت})}$$

(حسب (6-3)).

حيث S الانحراف المعياري للعينة المسحوبة من ذلك المجتمع و $\overline{X}$ متوسطها.

ثم نحسب قيمة إحصاء الاختبار عندما تكون H_0 صحيحة أي:

$$T_0=\frac{X-\mu_0}{\frac{S}{\sqrt{n}}}$$

واعتمادا على مستوى الدلالة α ونوع الاختبار ونوع التوزيع (ستودنت بـ $v=n-1$ درجة من الحرية) نحدد مناطق الرفض والقبول (كما ورد في (6-11)) ونقارنها مع T_0 ونتخذ القرار المناسب.

مثال (6-15):

إذا كان متوسط الزمن اللازم لتصنيع قطعة ما 50 دقيقة بانحراف معياري 10 دقائق. وبعد اعتماد طريقة جديدة في التصنيع وتجريبها على عينة من 12. كان متوسط الزمن اللازم للتصنيع هو 42 دقيقة بانحراف معياري 11.9 دقيقة. هل هذا يعني أن متوسط الزمن اللازم للتصنيع للقطعة قد اختلف وذلك بمستوى $\alpha=0.05$ من الأهمية.

الحل:

الفرضية الابتدائية: $H_0 : \mu=50$

الفرضية البديلة: $H_1 : \mu\neq 50$

ومستوى الدلالة $\alpha=0.05$ والاختبار من الطرفين وإحصاء الاختبار يكون

$$T=\frac{X-\mu}{\frac{S}{\sqrt{n}}} \sim t_{(\nu=n-1)} \quad (\text{ستودنت})$$

ثم نحسب قيمة إحصاء الاختبار عندما تكون H_0 صحيحة.

$$T_0=\frac{X-\mu_0}{\frac{S}{\sqrt{n}}}=\frac{42-50}{\frac{11.9}{\sqrt{12}}}=-2.33$$

واعتمادا على مستوى الدلالة $\alpha=0.05$ والاختبار من الطرفين والتوزيع لستودنت حيث

$$\alpha=0.05 \Rightarrow \frac{\alpha}{2}=0.025 \ ; \nu=n-1=12-1=11$$

$$t_{\alpha/2}(\nu)=t_{0.025}^{(11)}=2.20$$

ومنه منطقة القبول:

$$\left[-t_{\alpha/2}, t_{\alpha/2}\right]=[-2.20 \ , 2.20]$$

ومنطقة الرفض:

$$\begin{cases} \left]-\infty, -t_{\alpha/2}\right[= \left]-\infty \ , -2.20\right[\\ \left]t_{\alpha/2} \ , +\infty\right[= \left]2.20 \ , +\infty\right[\end{cases}$$

وبمقارنة T_0 والتي تساوي (2.33-) مــع منــاطق الرفــض والقبــول نجــد أن $]-2.20\ ,\ -\infty[\ni T_0$ أي في منطقة الرفض من جهة اليسار وبالتالي نرفــض H_0 ونقبل H_1 أي أن متوسط الزمن اللازم لتصنيع القطعة باعتماد النظام الجديد قد تغــير عن النظام القديم. وكون الرفض من جهة اليسار فإن الزمن اللازم للتصنيع قد انخفض عن الزمن القديم.

16-6: اختبارات حول الفرق بين متوسطين مجتمعين إحصائيين يتوزعان على وجه التقريب طبيعيا وبعينات صغيرة الحجم:

بالعودة لمعطيات الفقرة (6-6) نرغب في اختبــار $H_0: \mu_1-\mu_2=d$ (حيــث d ثابث وعلـى الغــالب يكــون d= 0 "أي لا فــرق بــين المتوســطين") مقــابل $H_1: \mu_1-\mu_2>d$ أو $H_1: \mu_1-\mu_2<d$ أو $H_1: \mu_1-\mu_2\neq d$. وعنــد مســتوى الدلالة α يكون إحصاء الاختبار:

$$T=\frac{(X_1-X_2)-(\mu_1-\mu_2)}{S_P\cdot\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim t_{(\nu=n_1+n_2-2)}\ (\text{ستودنت})$$

حيث:

$$S_P=\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}$$

ثم نحسب إحصاء الاختبار عندما تكون H_0 صحيحة:

$$T_0=\frac{(X_1-X_2)-d}{S_P\cdot\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

ثم نحدد اعتمادا على قيمة α ونــوع الاختبــار والتوزيــع لســتودنت بــــ $\nu=n_1+n_2-1$ (درجة من الحرية).

مناطق الرفض والقبول ونقارنها مع T_0 ونتخذ القرار المناسب حينئذ (تماما كمــا ورد في (6-11)).

مثال: (16-6) :

اختيرت مجموعتان من الطلاب $n_1=12, n_2=10$ وطبقت عليها طريقتان مختلفتان في التعليم. وفي نهاية الفصل الدراسي أجري لها اختبار واحد. فكان متوسط درجات المجموعة الأولى 85 بانحراف معياري 4 درجات، وبينما سجلت المجموعة الثانية متوسطا قدره 81 بانحراف معياري قدره 5. هل طريقتا التعليم متكافئتان بمستوى $\alpha=0.10$ من الأهمية.

الحل:

ليكن μ_1 متوسط درجات الطلاب وفق الطريقة الأولى.

و μ_2 متوسط درجات الطلاب وفق الطريقة الثانية.

الفرضية الابتدائية: $H_0: \mu_1=\mu_2$

الفرضية البديلة $H_1: \mu_1 \neq \mu_2$

والاختبار من الطرفين بمستوى دلالة $\alpha=0.10$ وإحصاء الاختبار:

$$T=\frac{(X_1-X_2)-(\mu_1-\mu_2)}{S_P.\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t_{(v=n_1+n_2-2)}\ (\text{ستودنت})$$

حيث:

$$S_P=\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}=\sqrt{\frac{(11)(16)+(9)(25)}{12+10-2}}$$

$$=4.478$$

ثم نحسب قيمة إحصاء الاختبار عندما تكون H_0 صحيحة:

$$T_0=\frac{(85-81)}{(4.478)\sqrt{\frac{1}{12}+\frac{1}{10}}}=2.07$$

وعند مستوى الدلالة $\alpha=0.10$ والاختبار من الطرفين والتوزيع لستودنت بـــ $v=n_1+n_2-2=12+10-2=20$ درجة من الحرية، فتكون منطقة القبول:

$$\left[-t_{\alpha/2}, t_{\alpha/2}\right]=[-1.725, 1.725]$$

ومنطقة الرفض:

$$\begin{cases} \left]-\infty, -t_{\alpha/2}\right[= \left]-\infty, -1.725\right[\\ \left]t_{\alpha/2}, +\infty\right[= \left]1.725, +\infty\right[\end{cases}$$

وبمقارنة $T_0 = 2.07$ مع مناطق الرفض والقبول نجـــد أن $\left]1.725, +\infty\right[$ $T_0 \in$ أي أنها تقع في منطقة الرفض من جهة اليمين ، وبالتالي نرفض H_0 ونقبـــل H_1 أي أن طريقتي التعليم غير متكافئتين وبما أن الرفض من اليمين فهذا يعني أن $\mu_1 - \mu_2 > 0$ أي $\mu_1 > \mu_2$ أي أن الطريقة الأولى أفضل من الطريقة الثانية.

17-6: اختبار فرضيات حول تباين مجتمع إحصائي σ^2:

رأينا من الفقرة (8. 6) أنه من أجل مجتمع طبيعي توقعه μ وتباينـــه σ^2 وعينـــة مسحوبة منه من الحجم n ، أن الإحصاء $X^2 = \frac{(n-1)S^2}{\sigma^2}$ يتوزع وفق كاي مربـــع بدرجة من الحرية $v = n-1$. وهنا سنختبر الفرضية الابتدائية $H_0: \sigma^2 = \sigma_0^2$ مقـــابل الفرضية البديلة $H_1: \sigma^2 > \sigma_0^2$ أو $H_1: \sigma^2 < \sigma_0^2$ أو $H_1: \sigma^2 \neq \sigma_0^2$. وعند مســـتوى الدلالة α يكون إحصاء الاختبار:

$$X^2 = \frac{(n-1)S^2}{\sigma^2} \sim X^2_{(v=n-1)} \text{ (مربع - كاي)}$$

ثم نحسب قيمة إحصاء الاختبار عندما تكون H_0 صحيح

$$X_0^2 = \frac{(n-1)S^2}{\sigma^2}$$

وعند مستوى الدلالة α ونوع الاختبار والتوزيع كاي - مربـــع بـــــ $v = n-1$ درجة من الحرية تكون:

من أجل $H_0: \sigma^2 = \sigma_0^2$ مقابل $H_1: \sigma^2 > \sigma_0^2$

منطقة القبول: $\left[0, X_{\alpha}^{2(v)}\right]$

منطقة الرفض: $\left]X_{\alpha}^2(v), +\infty\right[$

ومن أجل $H_0: \sigma^2 = \sigma_0^2$ مقابل $H_0: \sigma^2 < \sigma_0^2$

تكون: منطقة القبول $\left[X_{1-\alpha}^{2(v)}, +\infty\right[$

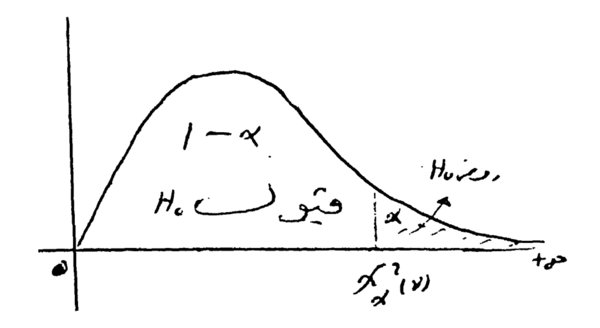
ومنطقة الرفض $\left[0, X^{2}_{1-\alpha}{}^{(\nu)}\right[$

ومن أجل $H_0 : \sigma^2 = \sigma_0^2$ مقابل $H_1 : \sigma^2 \neq \sigma_0^2$

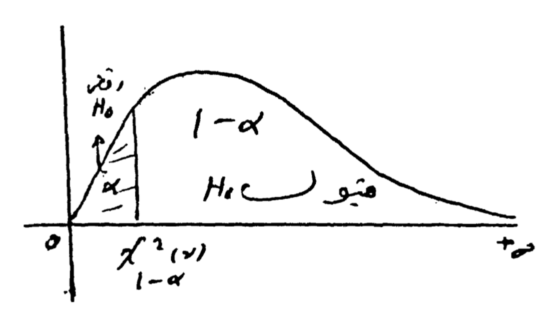
تكون منطقة القبول : $\left[X^{2}_{1-\alpha/2}{}^{(\nu)}, X^{2}_{\alpha/2}{}^{(\nu)}\right]$

ومنطقة الرفض: $\begin{cases} \left[0, X^{2}_{\alpha/2}{}^{(\nu)}\right[\\ \left]X^{2}_{\alpha/2}{}^{(\nu)}, +\infty\right[\end{cases}$

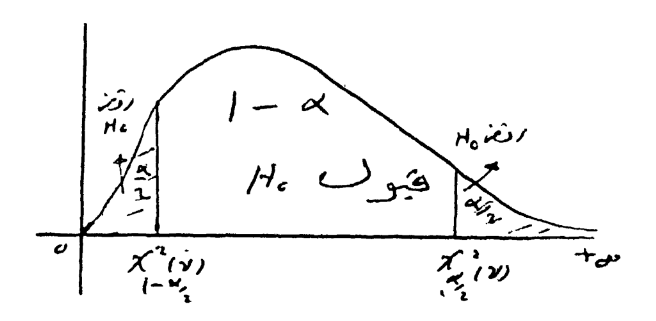
ثم نقارن مناطق الرفض والقبول مع X_0^2 (قيمة إحصاء الإختبار) ونتخذ القــــرار المناسب كما ورد في (6-15).



(اختبار من اليمن)



(اختبار من اليسار)



(اختبار من الطرفين)

ملاحظة :

يمكن إجراء الدراسة نفسها لاختبار فرضيات حول الانحراف المعياري σ مع الأخذ بالحسبان الجذر التربيعي الموجب.

مثال: 6-17:

يدعى مصنع لصنع المدخرات الكهربائية أن هذه المدخرات تعيش وسطيا ثلاث سنوات بانحراف معياري سنة واحدة. أخذت خمس مدخرات من إنتاج هذا المعمل فوجد أن أعمارها بالسنوات كما يلي:

1.9 ; 2.4; 3.0; 3.5; 4.2

هل نقبل بادعاء المصنع حول تباين عمر المدخرات بمستوى $\alpha=0.05$ من الأهمية، علما بأن عمر المدخرات يتوزع طبعيا.

الحل:

الفرضية الابتدائية $H_0: \sigma^2 = 1$

الفرضية البديلة $H_1: \sigma^2 \neq 1$

ومستوى الدلالة الإحصائية $\alpha=0.05$ والاختبار من الطرفين إن إحصاء الاختبار يكون

$$X^2 = \frac{(n-1)S^2}{\sigma^2} \sim X^2_{(v=n-1=5-1=4)} \quad (\text{مربع - كاي})$$

ثم نحسب قيمة إحصاء الاختبار عندما تكون H_0 صحيحة:

$$X_0^2 = \frac{(n-1)S^2}{\sigma_0^2} = \frac{(4)(0.815)}{1} = 3.26$$

حيث :

$$s^2 = \frac{n\sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i\right)^2}{n(n-1)} = \frac{5(48.26)-(15)^2}{(5)(4)} = 0.815$$

واعتمادا على مستوى الدلالة $\alpha=0.05$ والاختبار من الطرفين والتوزيع لكاي مربع بـ $v=4$ درجة من الحرية.

تكون منطقة القبول

$$\left[X^2_{1-\alpha/2}{}^{(v)}, X^2_{\alpha/2}{}^{(v)}\right] = [0.484, 11.143]$$

ومنطقة الرفض:

$$\begin{cases} \left[0, X^{2\,(\nu)}_{1-\alpha/2}\right[= [0, 0.484[\\ \left]X^{2\,(\nu)}_{\alpha/2}, +\infty\right[=]11.143, +\infty[\end{cases}$$

وبمقارنة $X_0^2 = 3.26$ مع مناطق الرفض والقبول نجد أن $X_0^2 \in [0.484,\ 11.143]$ أي تقع في منطقة القبول وبالتالي تقبل H_0 ونرفض H_1 أي أن إدعاء صاحب المصنع بأن تباين عمر المدخرات هو فعلا سنة واحدة.

18-6: اختبار فرضيات حول نسبة تباينين:

بالعودة إلى الفقرة (9.6) ومن أجل مجتمعين إحصائين توقع الأول μ_1 وتباينه σ_1^2 وتوقع التالي μ_2 وتباينه σ_2^2 ويتوزعان طبيعيا ومن أجل عينة عشوائية من المجتمع الأول من الحجم n_1 بتباين S_1^2 وعينة مستقلة عن الأولى من المجتمع الثاني من الحجم n_2 بتباين S_2^2. نرغب في اختبار الفرضية الابتدائية $H_0 : \sigma_1^2 = \sigma_1^2$ مقابل الفرضية البديلة: $H_1 : \sigma_1^2 > \sigma_1^2$ أو $H_1 : \sigma_1^2 < \sigma_1^2$ أو $H_1 : \sigma_1^2 \neq \sigma_1^2$.

عند مستوى الدلالة α

إن إحصاء الاختبار:

$$F = \frac{S_1^2}{S_2^2} \cdot \frac{\sigma_2^2}{\sigma_1^2} \sim f_{(\nu_1 = n_1 - 1,\, \nu_2 = n_2 - 1)}(\text{مثيشر})$$

ثم نحسب قيمة إحصاء الاختبار عندما تكون H_0 صحيحة فنجد عندئذ:

$$F_0 = \frac{S_1^2}{S_2^2}$$

واعتمادا على مستوى الدلالة α ونوع الاختبار والتوزيع لفيشر بدرجة من الحرية $(\nu_1 = n_1 - 1\ ,\ \nu_2 = n_2 - 1)$ نحدد مناطق الرفض والقبول:

فمن أجل : $H_0 : \sigma_1^2 = \sigma_2^2$ مقابل $H_1 : \sigma_1^2 > \sigma_2^2$

تكون منطقة القبول : $[0, f_\alpha(\nu_1, \nu_2)]$

ومنطقة الرفض: $]f_{\alpha}(\nu_1,\nu_2),+\infty[$

ومن أجل $H_0:\sigma_1^2=\sigma_2^2$ مقابل $H_1:\sigma_1^2<\sigma_2^2$ تكون:

منطقة القبول : $[f_{1-\alpha}(\nu_1,\nu_2),+\infty[$

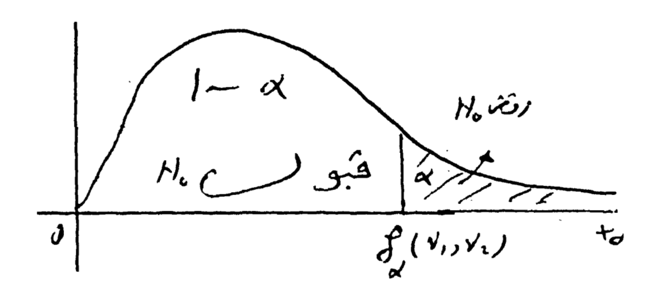
ومنطقة الرفض: $[0,f_{1-\alpha}(\nu_1,\nu_2)[$

ومن أجل : $H_0:\sigma_1^2=\sigma_2^2$ مقابل $H_1:\sigma_1^2\neq\sigma_2^2$ تكون

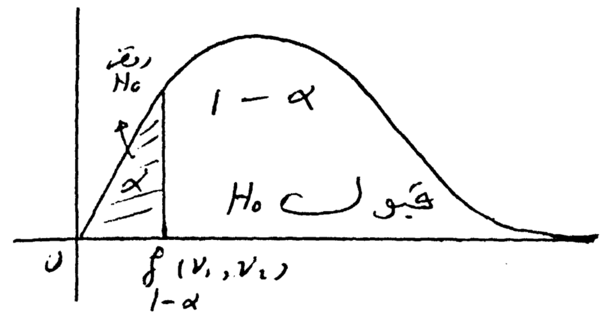
منطقة القبول : $\left[f_{1-\alpha/2}(\nu_1,\nu_2),f_{\alpha/2}(\nu_1,\nu_2)\right]$

ومنطقة الرفض: $\left[0\ ,f_{1-\alpha/2}(\nu_1,\nu_2)\right[\ ;\ \left]f_{\alpha/2}(\nu_1,\nu_2)\ ,+\infty\right[$

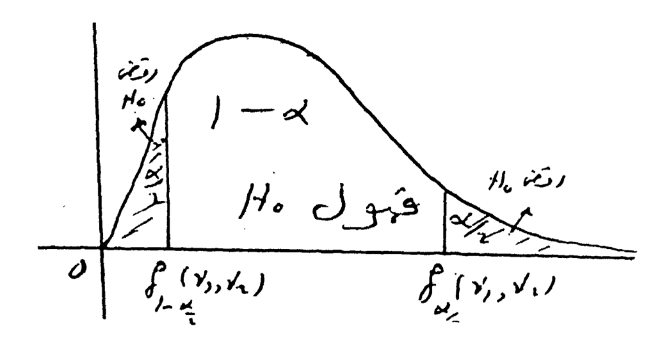
ثم نقارن F_0 مع مناطق الرفض والقبول ونتخذ القـــرار المناســب كمــا ورد في (11-6) .



(اختبار من الطرف الأيمن)



(اختبار من الطرف الأيسر)



(اختبار من الطرفين)

مثال: (6-18):

في اختبار لجودة قياس جهازي فولت من حيث الدقة لنوعين مختلفين B,A فـالنوع A أعطى من 8 قياسات تشتت 7.14 والنوع B أعطى من 10 قياسات تشتـــت 3.21 فهل تقدم هذه النتائج دلالة كافية على تساوي التشتتين عند المستوى $\alpha=0.10$ مـــن الأهمية. (علماً بأن المجتمعات المدروسة تتوزع طبيعياً).

الحل:

بفرض أن توزيعي المجتمعين يحققان شرط التوزيع الطبيعي بتوقع μ_1 ، μ_2 وتبـاين σ_1^2, σ_2^2 على الترتيب

فالفرضية الابتدائية $H_0 : \sigma_1^2 = \sigma_2^2$

والفرضية البديلة: $H_1 : \sigma_1^2 \neq \sigma_2^2$

ومستوى الدلالة $\alpha=0.10$ والاختيار من الطرفين

$$\alpha=0.10; \tfrac{\alpha}{2}=0.05; \nu_1 = n_1 - 1 = 8 - 1 = 7$$

$$\nu_2 = n_2 - 1 = 10 - 1 = 9$$

$$F_{\alpha/2}^{(\nu_1,\nu_2)} = F_{0.05}^{(7,9)} = 3.29$$

$$F_{1-\alpha/2}^{(\nu_1,\nu_2)} = F_{0.95}^{(7,9)} = \frac{1}{F_{0.05}^{(9,7)}} = \frac{1}{3.68} = 0.272$$

فمنطقة القبول: $\left[F_{1-\alpha/2}^{(\nu_1,\nu_2)}, F_{\alpha/2}^{(\nu_1,\nu_2)}\right] = [0.272, 3.29]$

ومنطقة الرفض: $[0, 0.272[,]3.29, +\infty[$

وإحصاء الاختبار

$$F = \frac{S_1^2}{S_2^2} \cdot \frac{\sigma_2^2}{\sigma_1^2} \sim f_{(\nu_1 = n_1 - 1, \nu_2 = n_2 - 1)}^{(\text{فيشر})}$$

ونحسب قيمته عندما تكون H_0 صحيحة:

$$F_0 = \frac{S_1^2}{S_2^2} = \frac{7.14}{3.21} = 2.22$$

ثم نقارن $F_0 = 2.22$ مع مناطق الرفض والقبول فنجد أن $F_0 \in = [0.27, 3.9]$ أي تقع في منطقة القبول ومنه نقبل H_0 ونرفض H_1 أي لا اختلاف بين بتايني دقـــة القياس في الجهازين.

تمرين عام (6-19):

تدعى شركة أعلاف أنه بتطبيق نظام غذائي معين من صنعها ، على نوع معين من الحيوانات، يمكن أن يزيد الوزن بمقدار 65 وحدة وزن خلال الأسابيع الثلاثة الأولى من حياها. ولاختبار هذا الادعاء، تم إخضاع عينة من 12 حيواناً من النوع نفسه للنظام الغذائي المذكور خلال الأسابيع الثلاثة الأولى من حياها، وبعد ذلك تم قياس الزيادة في الوزن لهذه الحيوانات فكانت كما يلي:

55	62	54	58	65	64
60	62	59	67	62	61

(وحدة وزن) فإذا كانت المجتمعات المدروسة هي تقريباً طبيعية).

المطلوب:

1- عيّن تقديراً منصفاً للمتوسط الحقيقي للزيادة في الوزن وللتباين الحقيقي للزيادة في الوزن والانحراف المعياري للزيادة في الوزن لمجتمع الحيوانات المطبق عليها النظام الغذائي المقترح.

2- عيّن قيمة الخطأ الأعظمي المرتكب في تقديرنا للمتوسط الحقيقي للزيادة في الوزن وبثقة 95%.

3- عيّن مجال ثقة حول المتوسط الحقيقي للزيادة في الوزن وبثقة 95%.

4- ما هو حجم العينة المناسب لتقدير المتوسط الحقيقي للزيادة في الوزن وبثقه 95% وبخطأ لا يتجاوز وحدة وزن واحدة.

5- هل تقدم المعلومات الواردة في التجربة المذكورة دليلاً كافياً على أن نظام التغذية الجديد يؤكد إدعاء الشركة أم لا، وذاك عند مستوى الأهمية $\alpha=0.05$.

6- لجأت شركة الأعلاف لتعديل النظام الغذائي المستخدم وذلك بإدخال بعض المركبات المغذية، ثم ادعت بأن النظام المعدّل يزيد من وزن الحيوان بمقدار 70 وحدة وزن خلال الأسابيع الثلاثة الأولى من ولادها، ولتأكيد ذلك جُرِبَ هذا النظام على

15 حيوانا ومن ثم قيست الزيادة في الوزن بعد ثلاثة أسابيع من حياتها وكانت كمايلي:

65	70	72	73	74
71	60	62	60	75
68	67	62	65	58

(وحدة وزن) وبفرض أن المجتمعات المدروسة تتوزع طبيعيا وتباينات متساوية.

a- هل تقدم هذه المعلومات دلالة كافية على أن نظام التغذية المعدل يؤكد إدعاء الشركة في مقدار زيادة الوزن وذلك عند مستوى الأهمية $\alpha=0.05$.

b- هل هناك فرق حقيقي بين متوسطي زيادة الوزن عند مجتمعي النظامين الغذائيين القديم والمعدل بأمثال ثقة 95%.

c- اختبر فرضية تساوي متوسطات الزيادة في الوزن بين مجتمع النظام الغذائي القديم ومجتمع النظام الغذائي المعدل ، وذلك عند مستوى الدلالة $\alpha=0.05$.

7- a– عين 95% مجال ثقة للتباين الحقيقي للزيادة في الوزن لمجتمع النظام الغذائي القديم.

b- عين 95% مجال ثقة للتباين الحقيقي للزيادة في الوزن لمجتمع النظام الغذائي المعدل.

8– a- ادعت شركة الأعلاف لأن التباين الحقيقي لتزايد الوزن عند استخدام النظام القديم هو 25 وحدة مربعة ، فهل النتائج التي حصلنا عليها من العينة المستخدمة تؤكد هذا الادعاء وذلك بمستوى $\alpha=0.05$ كم الأهمية.

b- وإذا ادعت الشركة بأن التباين الحقيقي لتزايد الوزن عند استخدام النظام المعدل هو 20 وحدة وزن مربعة. فهل النتائج التي حصلنا عليها من العينة المستخدمة بتطبيق النظام المعدل تؤكد ادعاء الشركة بمستوى $\alpha=0.05$ من الأهمية.

9- عين 90% مجال ثقة من أجل نسبة التباينين الحقيقي لمجتمعي العينين المدروستين والمتعلقتين الأولى بالنظام الغذائي القديم والثانية بالنظام الغذائي المعدل. وماذا نستنتج؟

10- اختبر فرضية تساوي التباينين الحقيقيين للزيادة في الوزن لمجتمعـــي العينتـــين المستخدمتين في المقارنة بمستوى أهمية $\alpha=0.10$ من أهمية.

الحل:

1- ليكن μ_1 المتوسط الحقيقي للزيادة في الوزن المتعلقة بالنظام الغذائي القديم و σ_1^2 التباين الحقيقي للزيادة في الوزن والمتعلقة بالنظام الغذائي القديم.

عندئذ:

$$\hat{\mu}_1=\overline{X}_1=\frac{\sum_{i=1}^{n}X_i}{n}=\frac{55+....+61}{12}=60.75$$

$$\hat{\sigma}_1^2=S_1^2=\frac{n\sum_{i=1}^{n}X_i^2-\left(\sum_{i=1}^{n}X_i\right)^2}{n(n-1)}=\frac{12(44449)-(729)^2}{12(11)}$$

$$\Rightarrow\hat{\sigma}_1^2=S_1^2 \qquad =14.75\Rightarrow S_1=3.8=\hat{\sigma}_1$$

2- إن حجم الخطأ الأعظمي المرتكب في تقديرنا لــ μ_1 وبثقة 95% يكون

$$e=\left|\pm t_{\alpha/2}(v).\frac{S_1}{\sqrt{n_1}}\right|.$$

$$v_1=n_1-1=12-1=11;\frac{\alpha}{2}=0.025;t_{\alpha/2}(v)=2.2$$

ومنه :

$$e_1=\left|\pm(2.2).\frac{3.8}{\sqrt{12}}\right|=2.42$$

3- إن $1-\alpha=0.9$ مجال ثقة حول μ_1 يكون من الشكل

$$X_1-e_1\leq\mu_1\leq X_1+e_1$$

$$60.75-2.42\leq\mu_1\leq60.75+2.42$$

$$58.33\leq\mu_1\leq63.17$$

ومن نكون 95% واثقين من أن μ_1 متوسط الزيادة في الوزن لن يقل عـــن 58.33 ولن يزيد على 63.17.

-4 $n=\left[\frac{t_{\alpha/2}(\nu)S_1}{e_1}\right]^2=\left[\frac{(2.2)(3.8)}{1}\right]^2=69.9\neq 70$

5- الفرضية الابتدائية: $H_0:\mu_1=65$

الفرضية البديلة: $H_1:\mu_1\neq 65$

ومستوى الدلالة $\alpha=0.05$ والاختبار من الطرفين

$$t_{\alpha/2}(\nu_1)=t_{0.025}(11)=2.2;\nu_1=n_1-1=12-1=1$$

ومنطقة القبول

$$\left[-t_{\alpha/2}(\nu_1),t_{\alpha/2}(\nu_1)\right]=[-2.2,2.2]$$

ومنطقة الرفض:

$$\begin{cases}\left]-\infty,-t_{\alpha/2}(\nu_1)\right[=\left]-\infty,-2.2\right[\\ \left]t_{\alpha/2}(\nu_1),+\infty\right[=\left]2.2,+\infty\right[\end{cases}$$

وإحصاء الاختبار

$$T_1=\frac{X_1-\mu_1}{\frac{S_1}{\sqrt{n}}}\sim t^{\text{ستودنت}}_{(\nu_1=n_1-1)}$$

وقيمة إحصاء الاختبار عندما تكون H_0 صحيحة

$$T_0=\frac{X_1-\mu_0}{\frac{S_1}{\sqrt{n}}}=\frac{60.75-65}{3.8}=-3.83$$

وبمقارنة T_0 مع مناطق الرفض والقبول نجد أن

$T_0\in\left]-\infty,-2.2\right[$ أي في منطقة الرفض من جهة اليسار ومنه نرفض H_0 وتقبل H_1 وبالتالي لا يمكننا تأكيد ادعاء الشركة بأن معدل الزيادة في الوزن هو 65.

6: a- لدينا

$$X_2=66.8;S_2^2=31.17,S_2=5.58$$
$$n_2=15;\quad \alpha=0.05$$

الفرضية الابتدائية $H_0:\mu_2=70$

الفرضية البديلة $H_1: \mu_2 \neq 70$

إحصاء الاختبار

$$T=\frac{X_2-\mu_2}{\frac{S_2}{\sqrt{n}}} \sim t^{\text{ستودنت}}_{(\nu_2=n_2-1=15-1=14)}$$

وقيمة هذا الإحصاء عندما تكون H_0 صحيحة:

$$T_0=\frac{66.8-70}{\frac{5.58}{\sqrt{15}}}=-2.221$$

وحسب $\alpha=0.05$ والاختبار من الطرفين والتوزيع لستودنت $\nu_2=14$ درجة من الحرية

$$t_{\alpha/2}(\nu_2)=t_{0.025}(14)=2.14$$

ومنطقة القبول:

$$\left[-t_{\alpha/2}(\nu_2), t_{\alpha/2}(\nu)\right]=[-2.14, 2.14]$$

ومنطقة الرفض:

$$\begin{cases} \left]-\infty, -t_{\alpha/2}(\nu_2)\right[=]-\infty, -2.14[\\ \left]t_{\alpha/2}(\nu_2), +\infty\right[=]2.14, +\infty[\end{cases}$$

وبمقارنة $T_0=2.221$ مع مناطق الرفض والقبول نجـــد أن $T_0 \in]-\infty, -2.14[$ تقع في منطقة الرفض من جهة اليسار وبالتـــالي نرفـــض H_0 ونقبـــل H_1 أي أن المعلومات في العينة لا تؤكد إدعاء الشركة بأن النظام المعدل يؤدي إلى زيادة في الوزن إلى 70 وحدة وزن لا بل أقل من ذلك كون الرفض من اليسار.

b-6 ليكن μ_1 متوسط الزيادة في الوزن في مجتمع الحيوانات وفق النظام القديم و μ_2 متوسط الزيادة في الوزن في مجتمع الحيوانات وفق النظام المعدل ولننشئ مجال ثقة 95% حول $\mu_1-\mu_2$: حيث لدينا المعلومات التالية:

النظام القديم $n_1=12$; $\overline{X}_1=60.75$; $S_1^2=14.75$; $S_1=3.8$

النظام الجديد	$n_2=15$; $\bar{X}_2=66.8$; $S_2^2=31.17$; $S_2=5.58$

$$t_{\alpha/2}(\nu)=t_{0.025}^{(n_1+n_2-2)}=t_{0.025}^{(25)}=2.06$$

$$S_P^2=\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}=\frac{(11)(14.75)+(14)(31.17)}{25}$$

$$=23.9452\Rightarrow S_P=4.89$$

$$t_{\alpha/2}(\nu).S_P.\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}=(2.06)(4.89).\sqrt{\frac{1}{12}+\frac{1}{15}}$$

$$=3.9$$

وإن $(1-\alpha)=0.95$ مجال ثقة حول $\mu_1-\mu_2$ يكون من الشكل التالي:

$$(\bar{X}_1-\bar{X}_2)-t_{\alpha/2}(\nu).S_P.\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\le\mu_1-\mu_2\le(\bar{X}_1-\bar{X}_2)+t_{\alpha/2}(\nu).S_P.\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

$$(60.75-66.8)-3.9\le\mu_1-\mu_2\le(60.75-66.8)+3.9$$

$$-9.95\le\mu_1-\mu_2\le-2.15$$

وبما أن طرفي المجال سالبان عندئذ $\mu_1<\mu_2\Rightarrow\mu_1-\mu_2<0$ أي أن معدل الزيـــادة في الوزن في اتباع النظام الغذائي المعدل أكبر منه في اتباع النظام القديم وذلك بثقة 95%.

c-6: الفرضية الابتدائية : $H_0:\mu_1=\mu_2$

الفرضية البديلة: $H_1:\mu_1\neq\mu_2$

ومستوى الدلالة $\alpha=0.05$ والاختبار من الطرفين وإحصاء الاختبار

$$T=\frac{(X_1-X_2)-(\mu_1-\mu_2)}{S_P.\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim t_{(n_1+n_2-2)}$$ (ستيودنت)

ونحسب قيمة هذا الإحصاء عندما تكون H_0 صحيحة:

$$T_0=\frac{(X_1-X_2)}{S_P.\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}=\frac{60.75-66.8}{(4.89)\sqrt{\frac{1}{12}+\frac{1}{15}}}=-3.20$$

وحسب $\alpha=0.05$ والاختبار من الطرفين والتوزيع لستودنت بـــ $\nu=25$ درجة من الحرية تكون منطقة القبول

$$\left[-t_{\alpha/2}(\nu_2), t_{\alpha/2}(\nu)\right]=[-2.06, 2.06]$$

ومنطقة الرفض:

$$\begin{cases} \left]-\infty, -t_{\alpha/2}(\nu)\right[=]-\infty, -2.06[\\ \left]t_{\alpha/2}(\nu), +\infty\right[=]2.06, +\infty[\end{cases}$$

وبمقارنة $T_0=-3.20$ مع مناطق الرفض والقبول نجـــد أن $T_0 \in]-\infty, -2.06[$ أي لمنطقة الرفض من جهة اليسار ومنه نرفض H_0 ونقبل H_1 أي أن النظامين غـــير متكافئين في معدل زيادة الوزن. وكون الرفض من جهة اليسار فإن $\mu_1-\mu_2<0$ أي $\mu_1<\mu_2$ وبالتالي النظام المعدل أفضل من النظام القديم في معدل زيادة الوزن.

7-a: إن $1-\alpha=0.95$ مجال ثقة حول σ_1^2 التباين الحقيقي لزيادة الـــوزن وفـــق النظام القديم:

$$\frac{(n_1-1)S_1^2}{X^2_{\alpha/2}(\nu_2)}<\sigma_1^2<\frac{(n_1-1)S_1^2}{X^2_{1-\alpha/2}(\nu_1)}$$

$$1-\alpha=0.95 \Rightarrow \tfrac{\alpha}{2}=0.025$$

$$\nu_1=n_1-1=12-1=11$$

$$X^2_{\alpha/2}(\nu_1)=X^2_{0.025}(11)=21.920$$

$$X^2_{1-\alpha/2}(\nu_2)=X^2_{0.975}(11)=3.816$$

ومنه يكون المجال:

$$\frac{(12-1)(14.75)}{21.920}<\sigma_1^2<\frac{(12-1)(14.75)}{3.816}$$

$$7.40<\sigma_1^2<42.52$$

$$2.72<\sigma_1<6.52$$

b-7: إن $1-\alpha=0.95$ مجال ثقة حول σ_2^2 تباين زيادة الوزن في المجتمع الثاني الموافق للنظام الغذائي المعدل:

$$\frac{(n_2-1)S_2^2}{X_{\alpha/2}^2(\nu_2)}<\sigma_2^2<\frac{(n_2-1)S_2^2}{X_{1-\alpha/2}^2(\nu_2)}$$

$$\frac{(15-1)(31.17)}{X_{0.025}^2(14)}<\sigma_2^2<\frac{(15-1)(31.17)}{X_{0.975}^2(14)}$$

$$\frac{(14)(31.17)}{26.119}<\sigma_2^2<\frac{(14)(31.17)}{5.629}$$

$$16.71<\sigma_2^2<77.52$$

$$4.09<\sigma_2^2<8.80$$

a-8: الفرضية الابتدائية : $H_0:\sigma_1^2=25$

الفرضية البديلة $H_1:\sigma_1^2\neq 25$

ومستوى الدلالة $\alpha=0.05$ والاختبار من الطرفين وإحصاء الاختبار

$$X^2=\frac{(n_1-1)S_1^2}{\sigma_1^2}\sim X_{(\nu_1=n_1-1)}^2 \quad (\text{كايمربع})$$

$$\nu_1=n_1-1=12-1=11$$

وقيمة إحصاء الاختبار عندما تكون H_0 صحيحة:

$$X_0^2=\frac{(12-1)(14.75)}{25}=6.49$$

واعتمادا على مستوى الدلالة $\alpha=0.05$ والاختبار من الطرفين والتوزيع لكاي مربع بـــ $\nu_1=11$ درجة من الحرية تكون منطقة القبول

$$\left[X_{1-\alpha/2}^2(\nu_1),X_{\alpha/2}^2(\nu_1)\right]=\left[X_{0.975}^2(11),X_{0.025}^2(11)\right]=[3.816,21.920]$$

ومنطقة الرفض:

$$[0,3.816[,]21.920,+\infty[$$

ثم نقارن $X_0^2=6.49$ مع مناطق الرفض والقبول فنجد أن

$X_0^2\in[3.816,21.920]$ أي تنتمي إلى منطقة القبول وبالتالي نقبل H_0 أي نقبل ادعاء الشركة بأن تباين زيادة الوزن باتباع النظام القديم هو 25.

b-8: الفرضية الابتدائية: $H_0:\sigma_2^2=20$

الفرضية البديلة: $H_1: \sigma_2^2 \neq 20$

ومستوى الدلالة $\alpha=0.05$ والاختبار من الطرفين وإحصاء الاختبار

$$X^2 = \frac{(n_2-1)S_2^2}{\sigma_2^2} \sim X^2_{(\nu_2=n_2-1=15-1=14)} \text{(كاي- مربع)}$$

وقيمة هذا الإحصاء عندما تكون H_0 صحيحة:

$$X_0^2 = \frac{(15-1)(31.17)}{20} = 21.819$$

واعتمادا على مستوى الدلالة $\alpha=0.05$ والاختيار من الطرفين والتوزيع لكـاي - مربع بـ $\nu_2=14$ درجة من الحرية تكون منطقة القبول:

$$\left[X^2_{1-\alpha/2}(\nu_2), X^2_{\alpha/2}(\nu_2)\right] = \left[X^2_{0.975}(14), X^2_{0.025}(14)\right] = [5.629, 26.119]$$

ومنطقة الرفض:

$$]0, 5.629[,]26.119, +\infty[$$

وبمقارنة $X_0^2 = 21.819$ مع مناطق الرفض والقبول نجد أن

$X_0^2 \in [5.629, 26.829]$ وبالتالي X_0^2 تقع في منطقة القبول أي تقبل H_0

ونرفض H_1 ومنه فإن ادعاء الشركة صحيح بأن تباين الزيادة في الوزن هو 20 وذلك بثقة 95%.

9- إن $1-\alpha=0.90$ مجال ثقة حول $\frac{\sigma_1^2}{\sigma_2^2}$ يكون من الشكل:

$$\frac{S_1^2}{S_2^2} \cdot \frac{1}{f_{\alpha/2}(\nu_1,\nu_2)} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2} \cdot f_{\alpha/2}(\nu_2,\nu_1)$$

$$\nu_1 = n_1 - 1 = 12 - 1 = 11; \nu_2 = n_2 - 1 = 15 - 1 = 14$$

$$\alpha = 0.10; \frac{\alpha}{2} = 0.05$$

$$f_{\alpha/2}(\nu_1,\nu_2) = f_{0.025}(11,14) = 2.65$$

$$f_{\alpha/2}(\nu_2,\nu_1) = f_{0.025}(14,11) = 2.72$$

ومنه:

$$\frac{(14.75)}{(31.17).(2.65)} < \frac{\sigma_1^2}{\sigma_2^2} < \left(\frac{14.75}{31.17}\right).(2.72)$$

$$0.178 < \frac{\sigma_1^2}{\sigma_2^2} < 1.287$$

وبالتالي نجد أن النسبة تقع في مجال يحوي الواحد أي أنه بثقة 90% يمكن أن يكون

$$\sigma_1^2 = \sigma_2^2 \Leftarrow < \frac{\sigma_1^2}{\sigma_2^2} = 1$$

وبالتالي لا فرق بين تبايني زيادة الوزن في المجتمعين المتعلقين بالنظام الغذائي القديم والنظام الغذائي الجديد أي أن المجتمعين متجانسان.

10- الفرضية الابتدائية : $H_0 : \sigma_1^2 = \sigma_2^2$

الفرضية البديلة: $H_1 : \sigma_1^2 \neq \sigma_2^2$

ومستوى الدلالة $\alpha = 0.10$ والاختبار من الطرفين وإحصاء الاختبار يكون

$$F = \frac{S_1^2}{S_2^2} \cdot \frac{\sigma_2^2}{\sigma_1^2} \sim f_{(\nu_1 = n_1 - 1 = 11; \nu_2 = n_2 - 1 = 14)}^{(\text{فيشر})}$$

وقيمة إحصاء الاختبار عندما تكون H_0 صحيحة:

$$F_0 = \frac{S_1^2}{S_2^2} = \frac{14.75}{31.17} = 0.473$$

واعتمادا على مستوى الدلالة $\alpha = 0.10$ والاختبار من الطرفين والتوزيع لفيشـر بـ $(\nu_1 = 11 \ , \ \nu_2 = 14)$ درجة من الحرية تكون منطقة القبول:

$$\left[f_{1-\alpha/2}(\nu_1, \nu_2), f_{\alpha/2}(\nu_1, \nu_2)\right] = \left[f_{0.95}(11,14), f_{0.05}(11,14)\right]$$

$$= \left[\frac{1}{f_{\alpha/2}(\nu_1, \nu_2)}, f_{\alpha/2}(\nu_1, \nu_2)\right] = \left[\frac{1}{f_{0.05}(14,11)}, f_{0.05}(11,14)\right]$$

$$= \left[\frac{1}{2.72}, 2.65\right] = [0.368, 2.65]$$

ومنطقة الرفض:

$$[0, 0.368[\, , \,]2.65, +\infty[$$

وبمقارنة $F_0 = 0.473$ مع مناطق الرفض والقبول نجد أن $F_0 \in [0.368, 2.65]$ أي أن F_0 تنتمي إلى منطقة القبول وبالتالي نقبل H_0 أي نقبل بأن تباينات الزيادة مـــن الوزن وفق النظام القديم والنظام الغذائي الجديد متساوية وذلـــك بثقـــة 90%. أي أن المجتمعات المدروسة متجانسة.

20-6: تمارين غير محلولة:

تمرين (1):

لنفرض أننا نرغب في تقدير متوسط الإنتاج اليومي في شركة للصناعات الكيميائية. وقد سجلنا الإنتاج اليومي لفترة n=60 يوما. فكان متوسط الإنتاج 941 طنا بانحراف معياري 23 طنا. والمطلوب تقدير متوسط الإنتاج اليومي لهذه الشركة μ. ثم إنشاء 0.99 مجال ثقة حول μ.

تمرين (2):

نعلم أن عمر خلية معينة من دارة كهربائية يتبع توزيعا احتماليا ملتويا. اخترنا عينة عشوائية من 250 خلية فكان متوسط العمر فيها 840 ساعة بانحراف معياري 21.98 ساعة. عين 95% مجال ثقة تقريبي لمتوسط العمر في مجتمع إنتاج هذه الخلايا.

تمرين (3):

نريد تقدير μ متوسط الطول في إنتاج مصنع للبراغي في حدود خطأ لا يزيد على 0.5مم إلا باحتمال لا يتجاوز الـ 0.05 فكم يجب أن يكون حجم العينة علما بأننا نعرف من سجلات الإنتاج السابقة أن الانحراف المعياري للطول يساوي 1.2مم.

تمرين (4):

من بين 300 أسرة اخترناها في منطقة معينة وجدنا 123 أسرة تمتلك تلفنازا ملونا عين 95% مجال ثقة لنسبة الأسر التي تمتلك تلفازا ملونا في كل هذه المنطقة.

تمرين (5):

سحبت عينة عشوائية حجمها 100 من مجتمع طبيعي، فكان متوسطها 50 وانحرافها المعياري 8، وسحبت عينة عشوائية حجمها 400 من مجتمع آخر مستقل عن الأول، فكان متوسطها 40 وانحرافها المعياري 12. عين 96% مجال ثقة حول الفرق الحقيقي بين متوسطي المجتمعين المدروسين وماذا نستنتج؟

تمرين (6):

أجري استفتاء لسكان المدينة والريف المحيط بها لمعرفة رأيهم حول اقتراح إنشاء ميدان للسباق، فصوت 2400 من أصل 5000 لصالح الاقتراح من سكان المدينة. كما صوت 1200 من أصل 2000 لصالح الاقتراح من سكان الريف المحيط بها. عين 90% مجال للثقة للفرق الحقيقي للنسبتين السابقتين.

تمرين (7):

إذا كان التباين لأعمار عينة من 30 مصباحا كهربائيا هو 100 ساعة، عين مجال ثقة بمستوى 95% حول الانحراف المعياري للعمر في المجتمع المصابيح.

تمرين (8):

إذا كان متوسط العمر الإنتاجي لعينة من 100 مصباح كهربائي من إنتاج مصنع معين هو 1570 ساعة بانحراف معياري قدره 120 ساعة. وإذا كان μ هو متوسط العمر الإنتاجي لمجتمع المصابيح المنتجة من المصنع.

اختبر ما إذا كان μ هو فعلا أقل من 1600 ساعة عند مستوى الأهمية $\alpha = 0.01$.

تمرين (9):

تدعي شركة لصناعة الأدوية أن أحد أدويتها الخاصة بمعالجة التحسس يحدث استجابة خلال فترة قصيرة لـ 0.80 من المرضى، ولاختبار هذا الادعاء أخذت عينة من 150 مريضا ، فوجد أن 110 منهم قد حدثت لهم استجابة فعلا، خلال الفترة المفروضة لدى تناولهم الدواء. فهل تقبل بصحة ادعاء الشركة بمستوى $\alpha=0.01$ من الأهمية.

تمرين (10):

رام محترف أصاب الهدف 60 مرة من 100 مرة إطلاق. وفي المئة الثانية أصاب الهدف 70 مرة هل يمكننا القول إن رمايانه قد تحسنت عند مستوى الأهمية $\alpha=0.05$.

تمرين (11):

أجري استفتاء بين سكان مدينة والريف التابع لها حول إنتشاء مركز صحـــي في طرف المدينة. فصوت 120 من أصل 200 أسرة من سكان المدينة لصـــالح المشــروع، بينما صوت 240 أسرة من بين 500 أسرة من سكان الريف لصالح المشروع. فـــهل تستنتج أن نسبة المصوتين لصالح المشروع من سكان الدينة تفوق مثيلتها من ســـكان الريف، بمستوى $\alpha=0.025$ من الأهمية.

تمرين (12):

أعطي نوعان من الأدوية بهدف تخفيف الألم الحادث بعد العمليات الجراحية، فمن أصل 100 مريض أعطي لهم الدواء A أدعى 38 منهم أنه خفف الألم، بينما من أصـــل 120 مريض تناولوا الدواء B أدعى منهم 56 مريض أنه خفف الألم.

هل ثمة دلالة على وجود فرق بين الدوائين بمستوى $\alpha=0.05$ من الأهمية.

تمرين (13):

يدعى صاحب مصنع للمصابيح الكهربائية أن متوسط العمر الإنتاجي للمصـــابيح التي ينتجها 1000 ساعة، أخذت عينة عشوائية من إنتاج هذا المصنع حجمها 25، فتبين بعد الفحص أن متوسط عمرها الإنتاجي 994 ساعة، بانحراف معياري 30 ساعة. فهل نقبل هذا الادعاء بمستوى $\alpha=0.05$ من الأهمية.

تمرين (14):

إذا علمنا أن الانحراف المعياري لآلة تزين الطرود البريدية هو 7.1 غراما. ومن عينة عشوائية من 20 طردا، وجد أن الانحراف المعياري لوزنها هو 9.1. هل هذا يعني ازدياد في الانحراف المعياري للآلة التي تزن الطرود بمستوى $\alpha=0.05$ من الأهمية.

تمرين (15):

طلبت وزارة التعليم العالي شحن من الأجهزة الكهربائية، واشترطت أن لا تزيـــد نسبة القطع السيئة الصنع على 7%. وقد ورد إليها ثلاث شحنات من ثلاثة مصـــادر مختلفة، أخذ من كل منها عينة عشوائية من 100 جهاز وكانت نسبة القطع الســـيئة الصنع فيها على التوالي:

0.01 ; 0.09 ; 0.12 . وقد قررت الوزارة منح مكافأة للمعمل الذي يقدم الأفضل. فما هو القرار الواجب اتخاذه بالنسبة لكل شحنة إذا كان مستوى الدلالة α=0.05 من الأهمية.

تمرين (16):

سحبت عينة عشوائية من 81 عاملا من وزارة التربية، نوجد أن متوسط أجورهم الشهرية L.S 2500 بانحراف معياري L.S 180. عين مجال ثقة حول متوسط أجور موظفي وزارة التربية بأمثال ثقة 99%.

تمرين (17):

سحبت عينة عشوائية من طلاب كلية التربية مؤلفة من 80 طالبا وعينة عشوائية ثانية من طلاب كلية الآداب مؤلفة من 90 طالبا وأخذت درجات امتحان الفصل الأول لهم فكان: متوسط درجات طلاب كلية التربية في العينة 65 بانحراف معيلوي 7 ومتوسط العينة لطلاب كلية الآداب 60 بانحراف معياري 5. فهل تعتقد بوجود مزق جوهري بين مستوى الطلاب في الكليتين بمستوى 0.05.

تمرين (18):

من عينة مؤلفة من 36 مريضا مصابين بمرض الإيدز، كان متوسط العمر الذي قضوه بعد الإصابة بالمرض هو 2.6 وبانحراف معياري 0.3 سنة والمطلوب

1- تعيين 99% مجال ثقة حول المتوسط الحقيقي للعمر الذي يقضيه مريض الإيدز بعد الإصابة بالمرض.

2- عين حجم العينة اللازم سحبه لكي نقدر μ بثقة 99% وبخطأ لا يتجاوز 0.05 سنة.

تمرين (19):

تبين من عينة عشوائية حجمها 100 متوفىء أن متوسط العمر لهؤلاء هو 71.8 سنة بانحراف معياري قدره 8.9 سنة فهل يشير هذا إلى أن مستوى العمر الآن يختلف عن 70 سنة بمستوى 0.05 من الأهمية.

تمرين (20):

قمنا بمقارنة نوعين من إطارات السيارات باختبار عملي يتضمن عينة حجمها 100 إطار من كل نوع. وسجلنا عدد الأميال التي يخدمها الإطار حتى اهترائه وفقا لمقاييس محدد سلفا وكانت النتائج كما يلي:

النوع الأول	$n_1 = 100$	$\overline{X}_1 = 2640$	$S_1^2 = 144000$
النوع الثاني	$n_2 = 100$	$\overline{X}_2 = 2510$	$S_2^2 = 196000$

عين 99% مجال ثقة حول الفرق الحقيقي بين متوسطي العمر في هذين النوعين وماذا تستنتج.

تمرين (21):

في دراسة حول التخلف العقلي لدى حدثي الولادة والناتج عن عدوى وراثية جاءت عن طريق الصبغيات الأنثوية، وجد أن هناك 4 حالات من عينة من 150 طفلا حديث الولادة يعانون من هذا التخلف.

1- عين 95% مجال ثقة للنسبة الحقيقية للمتخلفين عقليا والناتج من العدوى الوراثية.

2- عين حجم العينة اللازم من أجل تقدير النسبة الحقيقية بثقة 95% وبخطأ لا يتجاوز 0.001.

تمرين (22):

من عينة من 800 شخص من المدينة A صوت لصالح المرشح C : (500) شخص ومن عينة من 600 شخص من المدينة 8 ، صوت لصالح المرشح C: 400 شخص. عين 99% مجال ثقة حول الفرق الحقيقي بين نسبة المقترعين لصالح المرشح C المدينتين وماذا تستنتج.

تم اختبر بمستوى $\alpha = 0.05$ وجود فرق جوهري بين نسبة المقترعين لصالح المرشحC.

تمرين (23):

أخذت عينة مؤلفة من 100 طالب من إحدى الجامعات ووجد أن 31 طالبا منهم قد نجحوا في صفهم. كما أخذت عينة مؤلفة من 50 طالبة، ووجد أن 10 طالبات

منههن قد نجحن في صفههن فهل يمكن أن نستنتج أن نسبة نجاح الطلاب في هـــذه الجامعة هي نفسها عند الطالبات وذلك بمستوى α=0.05 من الأهمية.

تمرين (24):

- إذا كانت أوزان سبع علب متماثلة لمادة غذائية (بالأونزات) كما يلي:

9.6 ; 10.2 ; 10.0 ; 10.4 ; 10.2 ; 0.8 ; 9.8

والمطلوب:

1- عين 95% مجال ثقة حول متوسط الوزن الحقيقي لعلب هذه المادة الغذائية علما بأن توزيع الوزن هو تقريبا طبيعيا.

2- عين حجم العينة اللازم لتقدير متوسط الوزن الحقيقي بثقـــة 95% وبخطـــأ لا يتجاوز 0.15 من الأونزة.

3- عين 95% حول التباين الحقيقي لوزن العلبة لهذه المادة الغذائية.

تمرين (25): يدعى معمل الأعلاف أن نظاما غذائيا لديه يؤدي إلى زيادة في الوزن بمعدل 800 غرام خلال الشهر الأول من ولادة الفروج. وللتأكد من ذلك أخذت عينة من 10 من الفراريج الحديثة الولادة وطبق عليها هذا النظام وكانت الزيادة في الـــوزن كالتالي:

950 ; 760 ; 790 ; 850 ; 810 ; 820 ; 800 ; 860 ; 795 ; 815

هل تؤكد هذه النتائج ادعاء الشركة بمستوى α=0.05 من الأهمية.

تمرين (26):

لتقدير تباين كمية النحاس المركز في نوع معين من النباتات والموجود على صفات أحد الأنهر اخترنا عشوائيا عينة مؤلفة من 16 بنية وحرقناها ثم حللنا الرماد الحـــاصل لكل بنته، فوجدنا كمية النحاس المركز كما يلي (حسب وحدة قياس معينة).

5	3	34	18	27	14	8	50
38	43	35	20	70	25	60	19

وبفرض أن المجتمع المدروس طبيعي يتوقع μ وتباين σ^2.

عين 90% مجال ثقة حول التباين الحقيقي σ^2 والدال على كمية النحاس المتواجدة في هذا النوع من النباتات.

تمرين (27):

لدى باحث القناعة بأن جهاز القياس الذي يستخدمه لديه يتغير معين بـــانحراف معياري 2. وقد سجل خلال تجربة القياسات التالية:

9.4, 8.1 , 10.2 , 5.2 , 4.1 , 6,2 , 7.3 , 6.5

فهل ت×غير هذه المقاسات قناعة الباحث في دقة القياس مـــن جهـــة انحرافـــه المعياري وذلك بمستوى $\alpha=0.05$ من الأهمية.

تمرين (28):

في عينة عشوائية مؤلفة من 25 فيرعا بالدم لديهم كما يلي: (بالثانية):

45	40	47	46	42	50	47	48
49	49	44	43	39	38	41	49
40	40	40	42	43	44	45	47
41							

1- عين 99% مجال ثقة حول الزمن الحقيقي لتختر الدم لدى سكان مدينة دمشق

2- عين 95% مجال ثقة حول التباين الحقيقي لزمن تخثر الدم لدى سكان مدينـــة دمشق.

3- إذا كان من المعلوم مسبقا أن زمن تخثر الدم لدى سكان مدينة دمشق هـــو 42 ثانية. فهل نتائج العينة تؤكد هذا الادعاء بمستوى من الأهمية $\alpha=0.05$.

4- إذا كان من المعلوم أن تباين زمن تخثر الدم لدى سكان مدينة دمشق هـــو 16 فهل تؤكد نتائج العينة هذا الادعاء بمستوى من الأهمية $\alpha=0.0$.

5- من عينة مؤلفة من 15 متبرعا بالدم من مدينة حمص، سجل زمن تخـــثر الـــدم لديهم فكان كما يلي (بالثانية):

42	43	39	39	40	45	46	44
40	40	41	48	46	40	40	

وإذا كان μ_1 متوسط زمن تخثر الدم لدى سكان مدينة دمشق

μ_2 متوسط زمن تخثر الدم لدى سكان مدينة حمص

σ_1^2 تباين زمن تخثر الدم لدى سكان مدينة دمشق

σ_2^2 متوسط زمن تخثر الدم لدى سكان مدينة حمص

a- عين 99% مجال ثقة حول $(\mu_1 - \mu_2)$ وماذا تستنتج.

b- عين 95% مجال ثقة حول $(\sigma_1 - \sigma_2)$ وماذا نستنتج.

C- اختبر بمستوى $\alpha=0.01$ فرضية تساوي μ_1 , μ_2.

d- اختبر بمستوى $\alpha=0.10$ فرضية تساوي σ_1^2, σ_2^2 .

تمرين (29):

لدينا عينتان كل منهما مؤلفة من 200 مصباح. فإذا كان متوسط العمر الإنتاجي والانحراف المعياري له من الصنف A وعلى الترتيب 242 ، 18 ساعة. أما العينة من الصنف B فكان متوسط العمر الإنتاجي والانحراف المعياري له: 228 ، 16 ساعة على الترتيب.

1- عين 95% مجال ثقة حول الفرق الحقيقي بين متوسطي العمر الإنتاجي لمجتمعين الصنفين A , B.

2- هل هناك فرق حقيقي بين المتوسطين بمستوى 0.01 من الأهمية.

تمرين (30):

من عينة مؤلفة من 400 شاب وواعي وعينة من 600 شاب مراهق ممن يشاهدون برنامجا تلفزيونيا معينا وجد أن 200 من الواعين و300 من المراهقين يؤيدون بشكل كبير فكرة البرنامج وموضوعاته. فهل يمكن أن نستنتج أن نسبة المؤيدين للبرنامج من الواعين تختلف عن نسبة المؤيدين للبرنامج من المراهقين. وذلك عند مستوى $\alpha=0.05$ و $\alpha=0.0$ من الأهمية.

ثم أنشئ 98% مجال ثقة حول الفرق الحقيقي بين النسبتين وماذا نستنتج؟

مصحح

المراجع العلمية

مراجع الكتاب

مراجع للاستزادة والاطلاع

المراجع العربية:

1- د. أنيس كنجو: 1980 - الإحصاء وطرق تطبيقه في ميادين البحث العلمي - مؤسسة الرسالة

2- د. أنيس كنجو: 1976 - الإحصاء الرياضي - جامعة دمشق

3- د. عدنان عموره: 1993 - الاحتمالات - جامعة دمشق

4- د. عدنان عموره: 1997 - تصميم التجارب - جامعة دمشق

5- د. محمد صبح: 1990 - مبادئ الإحصاء والاحتمال - جامعة دمشق

6- د. عزات قاسم: 1995 - مبادئ الإحصاء والاحتمال - جامعة دمشق

7- د. إبراهيم محمد العلي: 1980 - نظرية العينات - جامعة حلب

8- د. أنور اللحام - الأستاذ شفيق ياسين 1982 - مبادئ الإحصاء والاحتمال - جامعة دمشق.

المراجع الأجنبية:

1- Anderson T.W – 1984 – <<An Introduction to Multivariate Statistical Analysis>> J. Wiley – New York.

2- B.Gnedenko – 1988- <<The Theory of Probability>> - MiR; Russ

3- CoCHRAN .W.G- 1977- <<Sampling Techniques>> - J .Wiley of Sons-Newyork – USA

4- Charles. R.H- 1982; <<Fundamental Concepts in the Design of Experiment>> -H.R.W – Newyouk – USA.

5- Draper, N.R and H.Smitb – 1981;<<Apphed Regression Analysis>> - J.Wiley- Newyork – USA.

6- Freund's – walpole – 1987 - <<Mathematical statistics>> - PHI- Internatianal – USA.

7- Lawrence C.Hamilton – 1992 - <<Regression With Graplncs>>; B.C. California- USA.

8- MILLER and Freund'- 1994- <<Probability and Statistics for Engineers>> - PH>New Jersey- USA.

9- Myers, R.H 1986-<<Classical and Modern Regression with applications>> - D.Press- Boston – USA.

10- Richard.A. J.D.W. Wichern – 1992- <<Multivarite Statistical Analysis>> - P.H- New Jersey – USA.

المصطلحات العلمية

إنكليزي - عربي

Acceptance Sampling	معاينة القبول
Aligned Systematic Sample	عينة نمطية مصنعة
Analytical Survey	مسح تحليلي
Autocorrelated	ذاتية الترابط
Analyse of Variance	تحليل التباين
Arithmetic mean	متوسط حسابي
Associative	تجميعي
Axiom	موضوعه
Alternative hypothesis	الفروض البديلة
Analyse of Covariance	تحليل التغاير
Assumptions	فروض
Between groups	بين المجموعات
Bias	تحيز
Binary	ثنائي
Binomial	ذو حدين (حداني)
Best linear unbiased estimate	أفضل تقدير خطي غير منحاز

Binomial distribution	توزيع حداني
Bounderies of strata	حدود الطبقات
Bowl	كيس
Classifcation	تصنيف
Cluster	عنقودي
Cluster Sampling	المعاينة العنقودية
Coefficient	معامل
Confidence	الثقة
Confidence Coefficient	معامل الثقة
Confidence interval	مجال الثقة
Confidence limits	حدود الثقة
Consistency	تماسك
Correction	تصحيح
Correction Coefficient	معامل التصحيح
Correlation	ارتباط
Correlation Coefficient	معامل الارتباط
Cost	تكلفة
Central limit theorem	مبرهنة النهاية المركزية.
Central moment	عزم مركزي
Chi- Square	كاي - مربع
Combinations	توافيق
Conditional Probability	احتمال شرطي

Continous Distribution	توزيع مستمر
Continous Random Variable	متغير عشوائي مستمر
Constant	ثابت
Convergence	تقارب
Covariance	تغاير
Canonical form	الشكل القانوني
Coefficient of Variation	معامل الاختلاف
Coding	ترميز
Confounding	ادماج
Consistent estimater	التقدير المنسق
Critical Value	قيمة حرجة
Combined ratio estimate	التقدير النسبة المركب
Cost function	دالة تكلفة
Data	بيانات إحصائية
Degrees of freedom	درجات حرية
Descriptive Surveys	مسوح وصفية
Design effect (Deff)	أثر التصميم
Domains of Study	ميادين دراسة
Couple Sampling	معاينة مضاعفة
Design of Experiments	تصميم التجارب
Decomposition	تفكيك
Density	كثافة

Density function.	دالة كثافة
Dependant	مرتبطة
Effective	فعاّل
Element	عنصر
Error	الخطأ
Estimate	تقدير
Expectation	التوقع
Expected Value	القيمة المتوقعة
Equivalence relation	علاقة التكافؤ
Exponential distribution	توزيع أسي
Experiment	تجربة
Errors of measurement	أخطاء قياس
Eye Estimate	تقدير بالعين المجردة
Effect	تأثير
Experimental units	وحدات تجريبية
Factor analysis	التحليل العاملي
Factorial Experiments	التجارب العاملية
Forecasting	التنبؤ
Fractional replication	تكرار جزئي
Frame	إطار
Frequency	تكراري
Function	دالة (تابع)

Finite population Correction	تصحيح مجتمع منته
Fisher Distribution	توزيع فيشر
Gama Distribution	توزيع غاما
Game	لعبة – مباراة
Graph	بيان
Geographic Stratification	تقسيم جغرافي إلى طبقات
Goodness of fit	حسن الملائمة
Gross Errors	أخطاء فاحشة
Hierarchical Analysis	التحليل المتشعب
Homogeneous	متجانس
Hypergeometric distribution	توزيع فوق الهندسي
Hypothes	فرضية
Interval	مجال
Independent	مستقل
Independents Variables	متغيرات مستقلة
Inequality	متباينة
Infinite	لانهائي
Index of inconsistency	دليل عدم اتساق
Inflation factor	عامل تضخيم
Interpolation	استيفاء
Intracluster Correlation	ارتباط ما ضمن العنقود
Inverse sample	عينة عكسية

Item	مفردة
Joint Distrubution	توزيع مشترك
Jacknife method	طريقة مدية الجيب
Lattice designs	التصميمات الشبكة
Least sguares method	طريقة المربعات الصغرى
Linear model	نموذج خطي
Linear regression	انحدار خطي
Limits	حدود
Lattice Sampling	معاينة شبكية
Linear regression estimator	مقدر انحدار خطي
Loss Function	دالة خسارة
Linear Form	ترابط خطي
Mean	متوسط
Median	وسط
Marginal Probability	احتمال هامشي
Mode	منوال
Moments	عزوم
Multinomial Distribution	توزيع حدودي
Normal Distribution	توزيع طبيعي
Numerical function	دالة عددية
Numbers	أعداد
Neyman allocation	محاصة بيمانية

Nonce Verage	عدم تغطية
Non- normality	لا طبيعية
Nonresponce	غير مستجيب
Observation	ملاحظة
Odds numbers	أعداد فردية
Operator	مؤشر
Ordered statistics	إحصاءات مرتبه
Outcome	نتيجة
Optimum allocation	محاصة مثلى
Over estimate	تقدير بالزيادة
Observed Value	القيمة المشاهدة
Parameters	الوسطاء
Path analysis	تحليل المسار
Pooled Estimate	تقدير متجمع
Pooled Variance	تباين مجتمع
Probability level	مستوى احتمالي
Point estimate	تقدير نقطي
Power of test	قوة الاختبار
Principal Components analysis	تحليل المركبات الأساسية
Plan	تصميم (خطة)
Population	مجتمع
Possiple Samples	العينات الممكنة

Pre- test	اختبار سابق
Proportion	التوزيع الاحتمالي
Proportion	تناسب
Proportional	متناسب
Proportional allocation	محاصة متناسبة
Purposive	عمدي
Purposive sampling	معاينة عمدية
Percentage	نسبة مئوية
Periodic Variation	تغير دوري
Pilot survey	مسح استطلاعي
Poststratification	تقيم بعدي إلى طبقات
Precision	إحكام
Primary Sampling Unit	وحدة معاينة أوليه
Probability Proportional to size sampling	معاينة باحتمال متناسب مع الحجم.
Product estimator	مقدر جدائي
Proportion Estimate	تقدير نسبة
Porposive selection	اختبار هادف
Parameter space	فضاء الوسيط
Partition	تجزئة
Poisson distribution	توزيع بواسون
Positive	موجب
Population mean	متوسط مجتمع

Quadratic form	شكل تربيعي
Quetient	خارج قسمة
Quadratic confidence limits	حدود ثقة تربيعية
Qualitative characteristics	خواص نوعية
Questionnaire	استبيان
Quota sampling	معاينة بالحصة
Qualitative Variable	متغير نوعي
Quantitative variable	متغير كمي
Range	مدى
Random model	نموذج عشوائي
Random sample	عينة عشوائية
Randomization	العشوائية (التعشية)
Relative efficiency	الكفاية النسبية
Relative Information	المعلومات النسبية
Replication	تكرار
Response	استجابة
Restricted	مقيد
Response variable	متغير الاستجابة
Regression Coefficient	معامل الانحدار
Residuale	بواقي (رواسب)
Random Sampling	معاينة عشوائية
Replecated Sampling	المعاينة التبديلية

Replace ment	إعادة
Ration	نسبة
Ratio estimates	مقدرات نسبة
Randomized response method	طريقة استجابة معشاة
Random numbers	أعداد عشوائية
Rare Items	مفردات نادرة
Record checks	تدقيق سجلات
Reinterview	إعادة مقابلة
Relative Precision	دقة نسبية
Repeated measurements	قياسات متكررة
Repeated sampling	معاينة متكررة
Response deviation	ميدان استجابة
Response Variance	تباين استجابة
Risk function	دالة مخاطرة
Random Vector	شعاع عشوائي
Random experimeant	تجربة عشوائية
Rank	رتبة
Relative frequency	تردد نسبي
Set	مجموعة
Sequence	متتالية
Simple	عينة
Simple random Variable	متغير عشوائي بسيط

Simple mean	متوسط عينة
Simple function	دالة بسيطة
Standardized normal distribution	توزيع طبيعي معياري
Standard deviation	انحراف معياري
Standardized random variable	متغير عشوائي معياري
Statistic	إحصاء
Statistical Data	بيان إحصائي
Student distribution	توزيع ستودنت
Subset	مجموعة جزئية
Sum	مجموع
Sum of squares	مجموع مربعات
Supremum	حد أعلى
Symmetric	متناظر
Sampler	معاين
Sample survey	مسح عينة
Sampling	معاينة
Sampling Unit	وحدة معاينة
Sampling Without replacement	معاينة دون إعادة
Sampling With replacement	معاينة مع إعادة
Self -weighting estimate	تقدير ذاتي الترجيح
Sensitive question	سؤال حساس
Simple random Sampling	معاينة عشوائية بسيطة

Skewness	التواء
Standard error	خطأ معياري
Strata	طبقات
Stratification	تقسيم إلى طبقات
Stratified random sampling	معاينة عشوائية طبقية
Stratum	طبقة
Subpopulation	مجتمع جزئي
Subunit	وحدة جزئية
Systematic Sampling	معاينة نمطية
Tables of distribution	جداول التوزيع
Test	اختبار
Total	كلي
Transformation	تحويل
Triple	ثلاثية
Target population	المجتمع الهدف
Three stage Sampling	معاينة على ثلاث مراحل
Total response Variance	تباين استجابة كلي
Travel Costs	تكاليف السفر
Two- Phase Sampling	معاينة ثنائية الطور (مضاعفة)
Two- Stage sawpling	معاينة على مرحلتين
Table ofrandom numbers	جدول الأرقام العشوائية
The ratio estimate	التقدير النسبي

Tests of significance	اختبارات المعنوية
Tests of hypothesis	اختبار الفرضية
Time series analysis	تحليل السلاسل الزمنية
Treatments	معالجات
Type II of error	خطأ من النوع الأول
Type I of error	خطأ من النوع الثاني
Unbiased estimate	تقدير غير منحاز
Unrestricted	غير مقيد
Unit	وحدة
Unaligned systematic sampling	معاينة نمطية غير مصففة
Uses of sample surveys	استخدامات مسوح العينة
Uniform distribution	توزيع منتظم
Uniqueness	وحدانية
Variable	متغير (متحول)
Variance	تباين
Variance estimate	تقدير تباين
Waiting time	زمن الانتظار

فهرس المحتويات

الفصل الثاني: المعاينة العشوائية البسيطة: 53



الجمعية التعاونية للطباعة
بدمشق

صدر باشراف لجنة الانجاز
سعر المبيع للطالب (١١٠) ل.س